AD-A080 377 NAVAL WEAPONS CENTER CHINA LAKE CA F/G 5/10
OPERATOR DECISION-MAKING CHARACTERISTICS.(U)
NOV 79 M J BARNES
UNCLASSIFIED NWC-TP-6124 NL

1 OF 1
AD
A080377

END
DATE
FILMED
3-80
DDC

(U) *Operator Decision-Making Characteristics*, by Michael J. Barnes. China Lake, Calif., Naval Weapons Center, July 1979. 80 pp. (NWC TP 6124, publication UNCLASSIFIED.)

(U) The purpose of this report is to review the operator's decision-making characteristics as a first step in developing decision aids for the air-to-ground combat environment. Both formal and empirical models of decision-making are discussed. Four components of decision-making: information selection, probability estimation, worth assessment, and action selection are identified. Empirical results from laboratory and real-world studies are used to characterize operator performance for each of these components of decision-making.

~~(U)~~ The results indicate that the operator is best characterized in terms of empirical, rather than normative, models of decision-making. In particular, models based on heuristic processes seem to best characterize operator performance. Heuristic processes are rules used by the operator which simplify the decision situation and result in predictable errors. Such rules are used because of the operator's processing limitations and not necessarily because of his irrationality. Decision-aiding is suggested as one way to overcome these limitations and improve operator performance.

CONTENTS

INTRODUCTION

During combat, the fighter/attack aircrew is faced with a dynamic, complex, and highly stressful environment. Even in noncombat missions, the increasing complexity of avionic systems puts a high informational load on their operators. The use of decision aids is a possible way to reduce the operator's cognitive load while improving overall systems performance. Presently, however, no overall structure exists to analyze the aircrew's decision environment and to pinpoint possible areas where decision aids might be useful.

This report is a review of the literature on human decision-making with a focus on possible applications for decision-aiding on Navy attack aircraft. The purpose of the report is to isolate general decision-making characteristics of the operator for future comparison with decision situations identified as important for naval attack missions. The operator's decision-making characteristics can then be compared to ideal solutions generated by mathematical decision models. The difference between operator performance and ideal performance will form one basis for developing criteria for interfacing the operator with decision-aiding devices.

Figure 1 shows how this report will be combined with information from other sources in order to develop these criteria. Basically, the criteria will be based on three considerations: (1) how the operator performs, (2) what he does, and (3) what aids are, or will be, available. The latter two considerations are subjects of ongoing research.

OVERVIEW OF OPERATOR DECISION-MAKING CHARACTERISTICS

A number of excellent reviews of human decision-making have been published recently (Nickerson and Feehrer, 1975; Slovic and Lichenstein, 1971; Slovic, Fischhoff, and Lichtenstein, 1977). The purpose of the present review is not to supplant these reviews, but rather to outline the human operator's decision-making characteristics (ODC) in order to generate a detailed scheme of decision-making tasks. The scheme and review of ODC will complement each other, since together they will characterize both different decision tasks and the operator's performance at these tasks.

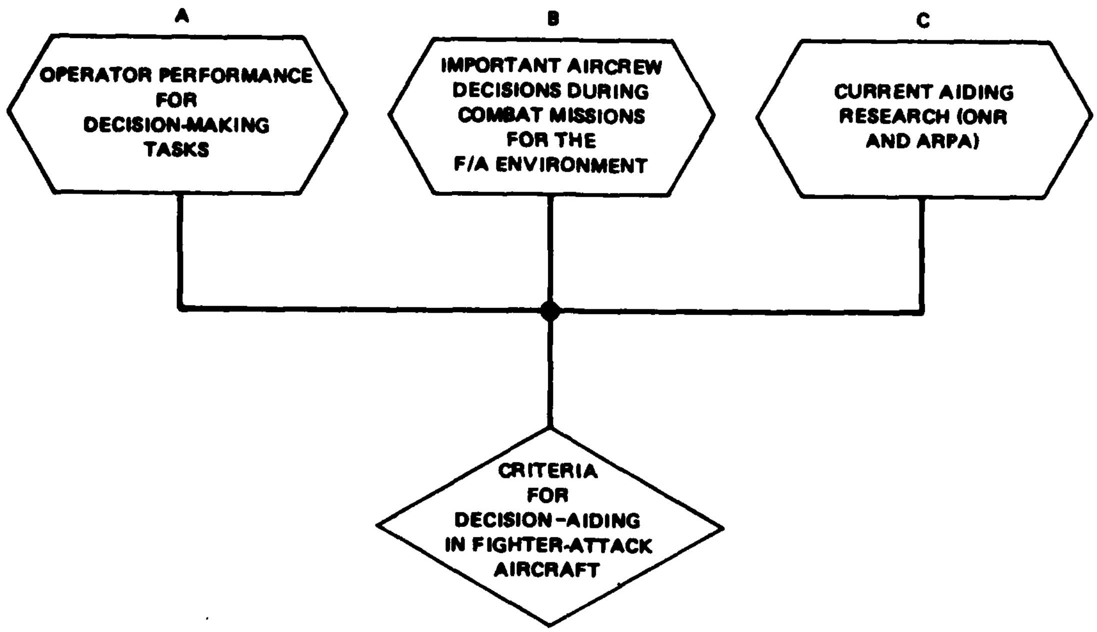


FIGURE 1. Development of Criteria for Decision-Aiding in Fighter/Attack Aircraft.

Figure 2 depicts a simplified model of probabilistic decision-making. It shows four stages of human decision-making: (1) information selection, (2) probability estimation, (3) worth assessment, and (4) action selection. The first stage consists of generating hypotheses concerning states of the world and selectively choosing data ($d_1 \ldots d_j \ldots d_n$) in the environment to test the hypotheses. Stage two involves probability estimation of different states of the world based on the data $[P(H_k|d_i) \ldots P(H_n|d_i)]$. Stage three involves generating possible outcomes ($\phi_1 \ldots \phi_n$) and evaluating their worth ($\phi_1 < \phi_j < \ldots < \phi_n$). The final stage is the consideration of alternative actions in light of the information from stages two and three.

An example may make the theoretical issues involved in decision-making more concrete. An attack pilot must decide whether to launch a missile at a ship for which he has only a noisy image on his CRT scope.

His first task is to decide how much information to collect before deciding whether to launch the missile or not. Too much information will result in putting himself in danger of being shot down and too little might result in his making the wrong decision. The operator's information selection characteristics will be discussed as an iterative process

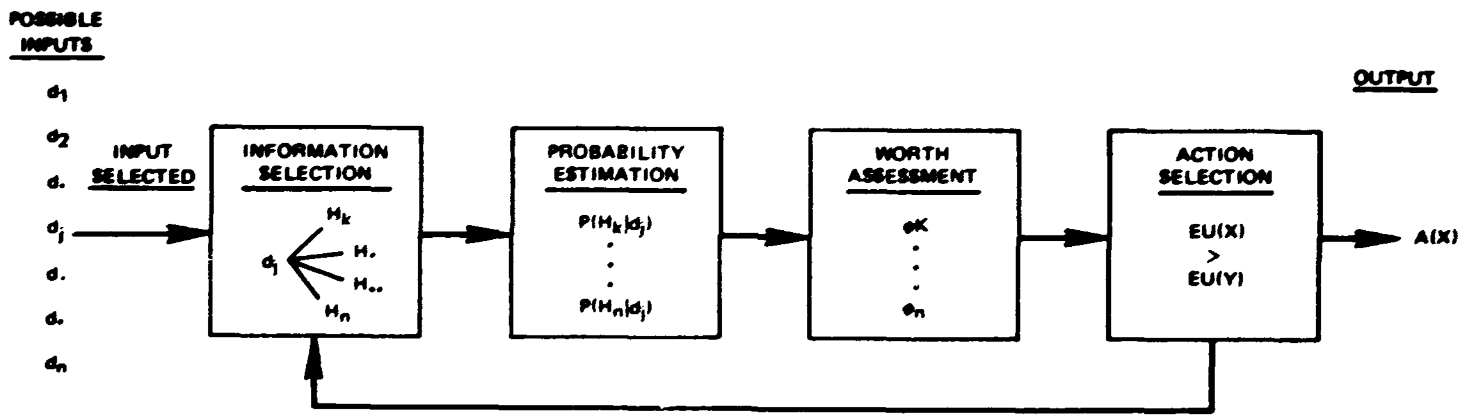


FIGURE 2. Stages of Probabilistic Decision-Making.

(c.f., the feedback loop in Figure 2), since the operator must know his state of uncertainty and the consequences of his possible actions before he can decide how much information to select. However, it is also the logical first step in decision-making since all the other stages require sufficient information in order to make even preliminary estimates.

After gathering information, the pilot can begin to form hypotheses about what he is viewing on his display. Based on both the data (d_j) and his prior experience, the pilot generates likely hypotheses. In our example, the imagery of the ship outline on his scope, especially the two faint superstructures, suggests an enemy destroyer or a friendly combatant (H_f or H_e). There is also a slight possibility that the image is a friendly hospital ship (H_h). Given the data (d_j) and anything else the operator knows about the situation, the probability estimation tasks consist of quantifying the operator's subjective level of uncertainty concerning each of the hypotheses [i.e., $P(H_e|d_j)$; $P(H_c|d_j)$ or $P(H_h|d_j)$].

However, his final decision depends not only on his subjective probability estimates, but also on his estimate of the consequences of his actions. The third stage of decision-making involves assessing the consequences of possible outcomes and is referred to as "worth assessment." In our example, the pilot must evaluate the possible outcomes of sinking or not sinking the ship imaged on his CRT. For example, the benefits of sinking an enemy ship [i.e., ϕ_e] might not balance the risk of sinking a friendly hospital [ϕ_h] or combat [ϕ_c] ship. The consequences of not sinking these ships must also be considered [e.g., $\phi_h = \phi_c < \phi_e$]. The discussion of worth assessment will concentrate on techniques for evaluating consequences and the performance characteristics of the operator using these techniques.

The final stage of decision-making consists of generating decision rules to select the best course of action using both the previously estimated probability and worth information. For example, because sinking a friendly may be much worse than to miss sinking an enemy ship, the operator may have a 90% criterion in order to launch the missile. If he is 90% sure that it is an enemy ship, he launches the missile; otherwise, he continues to track the ship. Thus the output in Figure 2 is a result of both the consequences of the action and the operator's state of subjective uncertainty as well as the particular decision rule he chooses.

This example will be used throughout the paper to introduce the different stages of decision-making. It would be misleading to assume that the example captures the difficulties inherent in real-world decision-making. In particular, the complexities associated with structuring the decision situation are, for the most part, unexplored. This is because little research has been done concerning how the operator structures his decision environment (e.g., deciding what options are realistic, what outcomes are possible...). Rather, the purpose of the example is to illustrate the issues relating to abstract decision models with an easily understood, concrete example.

The four-stage model of decision-making (Figure 2) will be the general framework for discussing features of various decision tasks and for characterizing the operator's performance. The functions in the boxes from Figure 2 can be interpreted in two ways. One, they can be normative functions following well-defined mathematical rules to optimize the expected outcome. Two, they can be descriptive functions characterizing what the operator actually does. Both interpretations are important in order to optimize performance in decision-making situations when a human operator is involved.

Figure 2 itself is too simplified, since it describes a static decision situation and most decisions are made in dynamic environments (cf. Rapaport, 1975). However, it does describe the main components of the decision-making process and seems to catch the essence of other decision-making models (Nickerson and Feehrer, 1973) while serving as a convenient framework for this report.

Probability estimation will be discussed first since this seems to be a necessary first step in decision-making for uncertain situations. Information selection will be discussed last since it involves concepts which will be developed in discussing other stages of decision-making.

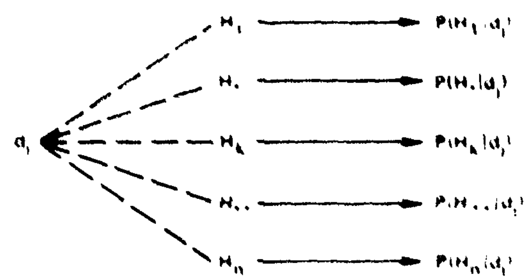
PROBABILITY ESTIMATION

OVERVIEW

Figure 3 illustrates the general framework of the probability estimation task. The operator must structure the problem by enumerating possible real-world states that are referred to as hypotheses ($H_1 \ldots H_k \ldots H_n$). His task is to express his degree of uncertainty regarding each H_k, based on its prior probability distribution ($P(H_k)$) and its relationship to incoming data ($d_1 \ldots d_j$). The quantification of the operator's subjective degree of uncertainty is his probability estimate.

For the example, the pilot would have two sources of information: (1) his subjective uncertainty concerning the image on the screen, and (2) a preflight intelligence briefing which indicated there was a 75% probability of an enemy ship being in the area and a 20% probability that a friendly destroyer might be in the area. The briefing also reported a remote possibility that a friendly hospital ship strayed off course and was in the area. The pilot's estimation task is to quantify the probabilities for each of the possible hypotheses.

Unfortunately, there are semantic and philosophic differences of opinion concerning the definition of probability. Presently, there are two main definitions: (1) the subjective (or personal) definition and (2) the frequency (or objective) definition (Hoel, 1971; Lee, 1971). The subjective definition (deFinetti, 1970; Savage, 1954) states that probability is the degree of uncertainty expressed behaviorally which follows certain mathematical axioms (e.g., the quantities sum to 1, are not negative, etc.). The frequency definition uses the same axioms, but defines probability as the expected distribution for repeated outcomes of the events being assigned probabilities. The main problem with the first definition is that the probability of an event can differ from operator to operator, whereas the second definition is not rich enough to allow probability estimates in a great number of cases (e.g., probability of a Russian attack tomorrow) (Edwards, Lindman and Savage, 1963; Savage, 1954).



WHERE $P(H_k|d_j)$ IS THE PROBABILITY OF HYPOTHESIS k GIVEN DATUM j

FIGURE 3. Outline of Probability Estimation Task.

The problem becomes even murkier because the operator's probability estimates do not necessarily follow the axioms that define probability and therefore are not subjective probabilities. In this report, the term subjective probability will be assumed to represent ideal operator performance and may or may not describe actual operator performance.

NORMATIVE MODELS OF ODC

Bayes Theorem

The traditional approach for investigating the operator's probability estimation characteristics was to compare his performance to 'ideal' mathematical models such as Bayes Theorem (Equation 1) (Peterson and Beach, 1967)

$$P(H_k|d_j) = \frac{P(H_k)\ P(d_j|H_k)}{\sum_{i=1}^{n} (H_i)\ P(d_j|H_i)} \tag{1}$$

where

$P(H_k|d_j)$ = probability of hypothesis k given datum j (i.e., posterior probability)

$P(H_k)$ = prior probability of hypothesis k

$P(d_j|H_k)$ = probability of sampling datum j given hypothesis k is true

$\sum_{i=1}^{n} (H_i)\ P(d_j|H_i)$ = probability of the data summed over all the possible hypotheses ($\Sigma\ P(H_i) = 1$)

Most traditional research involves binomial data, and it is convenient to compare the relative probability of two hypotheses (e.g., H_k and H_l) as an odds ratio.

$$\frac{P(H_j|d_j)}{P(H_l|d_j)} = \frac{P(H_k)}{P(H_l)} \cdot \frac{P(d_j|H_k)}{P(d_j|H_l)}\ , \quad \text{or} \tag{2}$$

$$\Omega_{k-1} = \Omega_o \cdot LK$$

where

Ω_{k-1} = posterior odds of hypotheses k to 1

Ω_{o} = prior odds

LK = likelihood ratio

Equation 3 shows the multiplicative effect of sampling *n* samples of conditionally independent data

$$\Omega_{k-1} = \Omega_{o} \prod_{1}^{n} LK \tag{3}$$

The revision of probability estimates as more data are sampled is referred to as Bayesian updating.

If we posit that the pilot in our example is a Bayesian estimator, his first task is to estimate his subjective probability of the particular image on his display, given the three hypotheses [e.g., $P(d_j|H_f) = .80$; $P(d_j|H_f) = .19$; $P(d_j|H_h) = .01$]. At this point, the operator may narrow his hypotheses to either a friendly or enemy ship, since the hospital hypothesis seems extremely unlikely. The Bayesian operator combines these estimates with the intelligence information [$P(H_f) = .75$; $P(H_e) = .20$] in order to compute the odds of the ship being enemy using Equation 2.

$$\Omega_{e-f} = \frac{.20}{.75}\,\frac{.80}{.19} = 1.12$$

Two points should be made concerning the above computation. (1) The operator state of uncertainty is high and thus his decision must rest mostly on the expected consequences of his actions. (2) His prior odds and likelihood information seem to cancel each other out. The latter point is important, since evidence will be introduced later which suggests that the operator tends to ignore the implications of prior odds.

At first glance, it seems a bit silly to argue that the operator uses Bayes theorem in a combat environment. However, this does serve as a useful first approximation model of the operator's probability estimation performance. By comparing the Bayesian estimate to the operator's actual estimate, it should be possible to see how well this model predicts operator performance. The initial research in this area

concluded that the operator was a good, if imperfect, "Bayesian" estimator (Peterson and Beach, 1967). It seemed as if the operator was estimating using something very like Bayes theorem, but due to his processing limitations was underestimating probability.

In order to understand this research, a sort of "generic" experimental paradigm is discussed next. Most of the research reviewed in this section (on normative models) used experimental paradigms very similar to the example to be presented, and one of the limitations to the research is the possible "artificiality" of these paradigms. The example will also help to illustrate the mechanics of the Bayesian updating task since the normative solution is Equation 3.

There are two urns (H_1 and H_2) from which the data (d_j) can be drawn. Urn 1 has 70% red balls (d_r) and 30% blue balls (d_b); the data proportions are the opposite for Urn 2. The experimenter chooses one urn at random (prior probability = .5) and draws 12 balls, one at a time, from the urn (replacing each ball before drawing another).

Given the fact that eight red balls and four blue balls were drawn, the subjects must estimate the probability (or odds) that the balls came from Urn 1. The subjects' estimated odds are almost always too conservative (i.e., they tend to underestimate the odds) for this type of task wherein the Bayesian solution (Equation 3) yields odds of 30 to 1 favoring hypothesis 1.

$$\Omega_{1-2} = \frac{(.5)}{(.5)} \cdot \frac{(0.7^8\ 0.3^3)}{(0.7^4\ 0.3^8)} = \left(\frac{0.7}{0.3}\right)^4 = 30$$

Figure 4 shows a conservative operator's probability estimates compared to the optimal solution and an excessive operator. The conservative operator underestimates the impact of data, whereas the excessive operator overvalues their impact. Notice that the more extreme the actual probabilities (abscissa), the more hypothetical operators deviate from the normative model. In general, it was found that the more diagnostic the data (i.e., the degree to which the data imply a particular hypothesis) the more subjects deviated from the Bayesian model (Peterson and Beach, 1967; Slovic and Lichenstein, 1971).

The almost universal finding of conservatism when using paradigms similar to our example was taken as evidence that the normative approach is fruitful. The operator apparently used some Bayesian method to update his probability estimates, but underestimated their magnitude. A correction for the slope values in Figure 4 should result in a conservative operator computing optimal solutions. Most of the normative research attempted to find the reason for suboptimal performance.

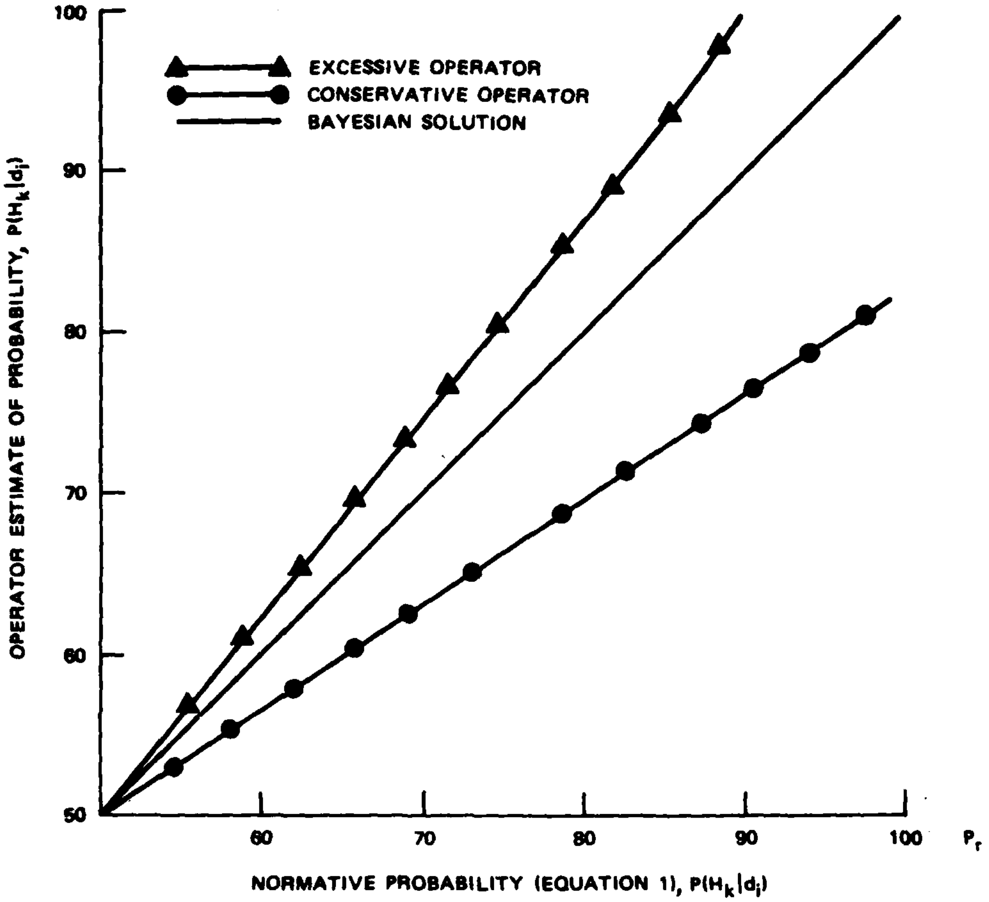


FIGURE 4. Hypothetical Characteristics of Operator in Bayesian Updating Task.

Empirical Results

The two main hypotheses for suboptimal performance were misperception and misaggregation. Misperception refers to the tendency of the operator to misperceive the properties of the underlying probability distribution. Peterson, Ducharme and Edwards (1968) found that subjects generated probability distributions that were too flat. If the probability estimates from these flat distributions were used with Equation 3, their subjects' likelihood ratios could be predicted.

Misaggregation refers to the inability of the operator to aggregate probability estimates optimally over a number of updating trials as implied by Equation 3. Ducharme and Peterson (1968) found that subjects' estimates of $P(d_j|H_k)$ for a single trial were nearly optimal but they could not properly aggregate the estimates over a number of trials. Apparently operators both misaggregate and misperceive the import of data in the updating task (Peterson and Beach, 1967; Slovic and Lichtenstein, 1971).

Besides misperceiving and misaggregating, the operator has other suboptimal estimation characteristics. Operators have a bias against giving extreme probability responses (Ducharme, 1970). Also, equal likelihood ratios (LK) are not interpreted as equal (e.g., $0.50/0.25 \neq 0.04/0.02$; Beach, 1968).

Furthermore, fallacious dependencies are seen to exist among independent events which cause operators to underestimate aggregated probabilities (Brickman and Pierce, 1972; Kahneman and Tversky, 1972). For example, Brickman and Pierce (1972) told their subjects the $P(d_j|H_k)$ for a single trial in an experiment similar to the Urn example. The subjects were told the color of balls drawn from Urn k for the past n trials and told to estimate the probability of a particular color ball being drawn on trial n + 1. If subjects were told that the preceding four balls were the same color, they underestimated $P(d_j|H_k)$ on trial five, because they assumed that drawing five balls of the same color was extremely unlikely. This tendency is well known in betting as "gambler's fallacy" (Lee, 1971).

Winkler and Murphy (1973) suggest that conservatism may be an artifact of the experimental paradigms used. These paradigms are highly artificial, whereas data in the real world are noisy and redundant (i.e., not independent), making it naturally nondiagnostic. If subjects treated data (non-noisy and independent) in these experiments like real-world data, they would be labeled conservative.

The manner in which probability estimates are made influences the subject's degree of conservatism. Stated probability estimates were more conservative than were probabilities inferred from overt acts (Godden, 1976). Since subjective probabilities technically are measured in terms of overt acts (such as betting, explicit preferences, etc.) the suboptimal characteristics may be partially an artifact of response mode (deFinetti, 1970).

Thus, there seems to be some doubt as to whether conservatism is the result of ODC or simply of the artificial paradigms used. Recent research, using both real-world data and different research paradigms, has found that operators tend to be excessive in as many situations as those in which they tend to be conservative (Fischhoff, Slovic, and Lichtenstein, 1977; Lichtenstein, Fischhoff, and Phillips, 1977; Pitz, 1975; Schaeffer, 1977; Slovic, 1972; Slovic, Fischhoff, and Lichtenstein, 1977). These later findings undermine the whole concept of the human operator being a suboptimal Bayesian performer. To quote Kahneman and Tversky (1972), the operator, "in his evaluation of evidence, is apparently not a conservative Bayesian; he is not Bayesian at all."

There has been a general shift in approach since 1972 from modeling the operator as Bayesian to attempting to find appropriate descriptive models for the ODC (Slovic, Fischhoff, and Lichtenstein, 1977).

DESCRIPTIVE MODELS OF ODC

Operator Bias and Heuristic Techniques

Most descriptive models imply that the human operator uses heuristic (i.e., simplifying) assumptions to generate probability estimates. A heuristic is a rule that simplifies computation, but does not always result in the correct conclusion. It may replace formal models (such as Equations 1, 2, and 3) and lead to predictable errors (biases). Different heuristics are used for different tasks. Tversky and Kahneman (1975) (also Kahneman and Tversky, 1972; Kahneman and Tversky, 1973) have identified the following three heuristics:

1. Representativeness: The use of salient properties of the parent population such as proportions, modal responses, etc., to predict sample outcomes. This heuristic involves the following biases: ignoring the implications of prior odds, sample size, randomness, and unreliability, and overweighting other features of the hypothetical population. This type of heuristic is almost deterministic. Some examples are the belief that long-haired students are radical, and that good law students necessarily become successful lawyers. Observed proportions with a small sample size are held to be as valid as same proportions with larger sample size.

2. Availability: The generation of frequency distributions by recalling examples of the relevant events. This heuristic involves biases associated with memory search. For example, vivid possibilities are judged more probable than less vivid ones.

3. Adjustment and Anchoring: A point in the distribution is used as a start and the distribution is adjusted to this anchor. Bias for this heuristic is related to the context of the available information. If subjects are given extreme information, their distribution will be adjusted to this information differently than if they are given median characteristics of the distribution. For example, subjects will construct different city population distributions if they are told that 95% of the cities in Brazil have a population under 25,000 than if they are told that the median city in Brazil has 5,000 people.

Slovic (1972) found that a subject's probability estimates depend on such factors as response compatibility and concreteness of the stimulus situation. The basic import of heuristic models is that the operator's strategy is more holistic than analytical. The operator will attempt to use any available information to improvise a simple strategy in order to make estimates.

A good example comes from research done by Marks and Clarkson (1972, 1973). For the Urn example given in the previous section, it can be shown that, since $P(d_R|H_1) = 1 - P(d_R|H_2)$, Equation 3 can be rewritten as:

$$\Omega_{1-k} = \Omega_{o} \, LK^{r-b} \tag{4}$$

where

r = number of red balls

b = number of blue balls

Although it simplifies computation, this relationship is not obvious. It implies that the posterior odds would be the same for samples of four red and one blue ball as for 54 red and 51 blue balls. This apparently is counter-intuitive, because performance of subjects in Marks and Clarkson's experiments (1972, 1973) was best predicted by Equation 5.

$$\Omega_{1-k} = \Omega_{o} \frac{r}{b} \tag{5}$$

Although Equation 5 has no basis in probability theory, 40% of the subjects reported using it. These subjects were using a simple strategy that takes advantage of representative information to estimate probabilities.

Statistics is a difficult and not always intuitive discipline. Tversky and Kahnemann (1975) point out that the intuition of even those researchers with years of experience in making statistical inferences often runs counter to normative prescriptions. It is not surprising, therefore, that the operator's intuition is often at odds with normative models.

Empirical Research

Most research dealing with operator biases and heuristics tend to concentrate on what psychological strategies the subject actually employs rather than to posit some suboptimal Bayesian process. Bayes theorem is a yardstick rather than a model of human performance for this type of approach.

Bayes theorem states that odds generated from the data must be weighted by the prior odds in order to compute the present probability. A good deal of research indicates that subjects ignore prior odds for a number of different experimental paradigms (Bar-Hillel, 1975; Lyon and Slovic, 1976; Nisbett and Borgida, 1975; Tversky and Kahneman, 1977).

The operator does not always understand how base rates (i.e., prior odds) affect probability estimates. For example, subjects were told that the base rate for cabs in a city was 85% blue and 15% green. Their task was to assess the probability that a witness who could tell blue from green cabs 80% of the time actually saw a green cab. Few subjects considered the effect of the different base frequency for blue and green cabs resulting in a modal response of 80% (which is the witness' hit rate) whereas the Bayesian solution was 41%. Changes in the base rate, cover story or hit rate failed to change these findings (Lyon and Slovic, 1977).

Tversky and Kahneman (1977) argue that the operator ignores base rate data when other data such as hit rate are seen to cause the phenomena being estimated. Thus, the subjects in Lyon and Slovic's experiment thought that the witnesses' hit rate was the cause of reporting the correct cab color and felt the relative frequency of the cabs was unimportant. Maya Bar-Hillel (1977) changed the story of the cabs so that hit rate was replaced by a random process (although in Bayesian terms the solution was the same). In this case, because there was no causal link between hit rate and the subject's report, many of the subjects considered base rate when making their estimates. The operator interprets the validity of different data sources in terms of their psychological significance rather in terms of their normative or Bayesian significance. (Bar-Hillel, 1977; Tversky and Kahneman, 1977).

The operator is also influenced by knowing the outcome when he assesses what his prior probabilities would have been. This is a fairly general phenomenon influencing both the operator's probability estimates and his memory of these estimates (Fischhoff, 1975, 1977). This "knew it all along" attitude causes the operator to underestimate his previous uncertainty once the true state of the world is revealed to him. Furthermore, the operator tends to be excessively confident concerning his knowledge of the world. Pitz (1975) found his subjects' confidence limits far too narrow when judging the populations of different cities. Fischhoff, Slovic and Lichtenstein (1977), using general knowledge questions, found their subjects' odds estimates to be extremely excessive. Lichtenstein and Fischhoff (1977) found their subjects were only moderately overconfident, but degree of overconfidence was not significantly decreased by comparing expert and non-expert subjects.

Shanteau (1972) using "functional measurement" techniques found evidence that his subjects were using an additive averaging rule and not the multiplicative Bayesian rule when updating their probability estimates. Eils, Seaver and Edwards (1977) found that allowing the subjects to update using a logarithmic response scale caused their subjects estimation performance to improve. Since Bayesian updating is additive on a logarithmic scale, this further suggests that some non-optimal additive averaging rule may be used as a heuristic for the updating task.

The research, taken together, belies the usefulness of assuming the operator is a Bayesian probability estimator. Rather, ODC seems to be task-specific. If the operator is asked to construct confidence limits on information about which he knows little, the limits will be excessive (Fischhoff, Slovic, and Lichtenstein, 1977; Lichtenstein, Fischhoff, and Phillips, 1977; Pitz, 1975). However, for tasks in which he must update probability information, the operator will usually be conservative (Peterson and Beach, 1967). The tasks reviewed so far involved either updating odds as new information is received or estimating the probability of an event on the basis of what the operator knows. The next sections involve more complex tasks and the final section reviews what we know about operator performance in real-world tasks.

COMPLEX ESTIMATION TASKS

Multistage Decisions

Multistage inferences involve data states that are removed from the hypotheses by one or more stages. The operator knows the relationship of a data set $(d_1 \ldots d_j \ldots d_n)$ to the hypothesis set (i.e., $P(d_j|H_k)$ is known). The operator also knows the relationship of a second data source $(r_1 \ldots r_g \ldots r_n)$ to the data set [e.g., if barometer 1 indicates 30 (r_1), the probability that the atmospheric pressure is actually 30 is 0.97 (i.e., $P(d_1|r_1)$]. The operator's task is to infer the posterior odds of H_k, if he is given r_i. The optimal solution is given by Equation 6.

$$\Omega_{k-1} = \frac{\sum P(r_i|H_k)\ P(d_j|H_k)}{\sum P(r_i|H_1)\ P(d_j|H_1)}\ \Omega_o \tag{6}$$

where $P(r_i)$ and $P(d_j|H_k)$ are conditionally independent (Schum and Pfeiffer, 1973).

Some examples of multistage inference problems are: attempting to compute the probability of real-world states $(H \ldots H_k \ldots H_n)$ if sensor reports $(r_i \ldots r_i \ldots r_n)$ are unreliable indicators of data states $(d_1 \ldots d_j \ldots d_n)$, reports from spies with differing credibilities, or a composite probability from witnesses who saw a crime.

Equation 7 is useful for multistage inference concerning two possible data states if only a general indication of the reliability (W) of the data source is known (Snapper and Fryback, 1971).

$$\Omega_{k-1} = \frac{P(d_j|H_k) + \mu}{P(d_j|H_1) + \mu} \cdot \Omega_o \tag{7}$$

where $\mu = \frac{1-W}{2W-1}$ and is undefined when $W = .5$.

Since neither equation is particularly intuitive, it is not surprising that subjects use heuristics in this task (Gettys, Kelly, and Peterson, 1973; Johnson et al, 1973; Snapper and Fryback, 1971). The "as-if" heuristic involves ignoring the probabilities of the initial data source (r_i) after choosing the most probable source. For example, a spy reports that enemy ships are leaving the port; this datum would be used to assess the probability of attack. The probability that the enemy ships did not leave port (i.e., spy is lying) is not considered. The operator acts "as-if" the report is perfectly reliable. This, of course, makes the estimates excessive.

For the "best guess" heuristic (Equation 8), the operator considers only the most probable report and weights the posterior odds with the reliability of this report (W_i).

$$\Omega_{k-1} = W_i \, LK \, \Omega_o \tag{8}$$

Equation 8 seems to predict operator performance best for a variety of multistage scenarios (Gettys, Kelly, and Peterson, 1973; Johnson et al, 1973; Snapper and Fryback, 1971).

There is evidence that the operator treats nonstationary probability situations like a multistage inference task. In nonstationary situations, the probability of a datum, given an hypothesis, is variable. For example, the probability that enemy ships leaving port implies an attack can change with a change in political climate. Operators apparently can use information that the underlying probability distribution is subject to change in the same manner as they use other multistage information (Chinnis and Peterson, 1968; 1970).

Conditional Non-Independence

Equation 3 implies that the data being sampled are conditionally independent, i.e., $P(d_1, d_2|H_k) = P(d_1|H_k)\ P(d_2|H_k)$. In the real world, some data are redundant $[P(d_1 \cap d_2) > Pd_1\ Pd_2]$, and some are particularly informative $[P(d_1 \cap d_2) < Pd_1\ Pd_2]$.

Lichtenstein (1972) found Equation 3 to be a robust predictor of disease type even when the data had considerable conditional non-independence. Since the type of non-independence was not discussed in detail, it is difficult to generalize these results to other situations.

Edwards (Beach, 1975; Domas and Peterson, 1972) suggests using a decision-aiding system he named PIP: the operator makes estimates of $P(d_j|H_k)$ and computers update these estimates using Equation 3. In general, this improves the efficiency of the updating task (Domas and Peterson, 1972; Schaefer, 1972). However, when the data presented to their subjects were redundant, Domas and Peterson (1972) found that PIP was less accurate than the posterior odds aggregated by their subjects. This supports the contention (Winkler and Murphy, 1973) that conservatism for the updating task may reflect the operator's experience with real-world data which are often redundant. There is evidence that the operators do take redundancy into account as well as make allowances for data which they feel are particularly informative (Wyer, 1970). However, considering the number of biases subjects show in the updating task, it seems unlikely that their estimates are more realistic than Bayesian ones. More information is needed on exactly how operators aggregate non-independent data. It seems probable that they use a simple heuristic rather than attempt to compute the often complex dependencies that can exist among data sources.

A simple optimal solution is given by Zlotnick (1968). He suggests combining different data sources into bundles which are themselves conditionally independent. Then, the operator (or a computer) can aggregate over bundles using Equation 3.

Probabilistic Model Generation

In many complex situations, the operator can use incoming data to predict the general trend of future events by constructing a model of his environment. The question addressed in this section is how well the operator can generate predictive models if the data are related to the general trend only probabilistically. The optimal strategy in such cases would be the least-square solution used in regression analysis.

For a medical diagnosis task, subjects were able to learn both linear and quadratic trends to a certain extent, but they did better at

predicting linear trends (Brehmer and Qvarström, 1976). Not surprisingly, the less variability in the incoming data, the better subjects were able to predict trends (Brehmer, Kuylenstierna, and Jergren, 1974). However, they were very poor at learning quadratic trends. They were even worse at using the trends they learned when they had to predict the correct solution. Again, quadratic trends were even more difficult to predict than were linear trends.

The reason for the latter results are related to some findings by Hammond (1972). There are two stages to conceptual development: acquisition of the concept, and cognitive control or output. For example, one might understand that a series of numbers increases quadratically and still do a poor job in generating the sequence. Hammond was able to demonstrate that cognitive control was more difficult than concept acquisition for a number of task situations. Also, Hammond, Stewart, Brehmer, and Steinmann (1975) found that subjects with the same basic cognitive model were in conflict because of their lack of output control.

Brehmer (1974) gives evidence which indicates that the operator uses a linear trend model as a first hypothesis. Only after negative results does the operator begin to entertain non-linear relationships to predict future trends. If a subject is trained to recognize non-linear trends, he is able to switch to linear situations rapidly. However, the reverse is not true. (Hammond et al, 1975.) This suggests that the operator normally entertains a very simplified model of the world. This model may be functional, simply because simple linear trends do a good job of predicting many types of complex trends, if there is a monotonic relationship among the predictor and criterion variables (i.e., X and Y).

There are a number of biases reported for probabilistic learning situations (Lee, 1971). The most prevalent is gambler's fallacy which has been reported previously. Another common bias is referred to as "probability matching." This bias involves events $[e_1 \ldots e_i \ldots e_n]$ occurring with specified probabilities $[p_1 \ldots p_i \ldots p_n]$. In the typical experimental paradigm, subjects must guess the next e_i for each trial for a large number of independent trials (they are usually told the correct e_i after each trial). The optimal strategy is to guess the single event with the largest probability for every trial. However, subjects usually guess every e_i with the same approximate relative frequency as its relative probability (i.e., match frequency of guessing e_i to p_i).

Most of the research reviewed indicates that the ability of the operator to predict future events based on uncertain cues is poor. However, Hammond et al (1975) reviewed research in which the subjects were quite good at making these types of predictions. This suggests that more research is necessary to find what task parameters are important determinants of the operator's ability to predict future trends.

REAL-WORLD TASKS

The purpose of the previous three sections was to characterize the operator in complex tasks which have some real-world validity. This section will report operator performance characteristics observed in actual real-world (non-laboratory) tasks.

Scoring Rules and Calibration

An important prerequisite for surveying real-world performance is to have some quantitative criteria for judging estimators. Two general approaches have been used to assess probability estimators.

1. Proper scoring rules are computational procedures which yield the maximum value to estimates that match observed frequency distributions. Linear payoff schemes (reward linearly related to number of correct outcomes) are not proper scoring rules. In this case, the operator can maximize his score by assigning 1.0 to the most probable category (Murphy and Winkler, 1971). Scoring rules using logarithmic, quadratic, and spherical functions have been shown to be proper (Murphy and Winkler, 1970). Although all these rules yield maximal values when predicted frequency matches observed frequency, some proper rules were more sensitive to incorrect estimates than others (Murphy and Winkler, 1971; Brown and Shuford, 1973). In particular, Hoffman and Peterson (1972) found a quadratic rule to have only minimal impact as feedback, presumably because it was insensitive to estimates that were not radically different from the correct estimate.

2. Another way to measure the operator's calibration (i.e., how close his estimates are to observed estimates) is to plot observed frequency of events against the operator's estimates (Lichtenstein, Fischhoff, and Phillips, 1977). This has the advantage of showing the direction as well as magnitude of operator errors (cf., Figure 2). As a convention, in this report calibration will refer to plotted data and scoring rules will be understood to refer to any of the proper scoring rules.

Operator Performance

Winkler and Murphy (1973) point out that probability estimation skill in the real world involves (1) substantive understanding of the events being predicted, and (2) normative understanding of the underlying probability structure. In other words, the estimator must know the dynamics of the real-world situation before he can relate them to a probability structure.

Weather forecasters must combine information of frequency trends with detailed knowledge of meteorological phenomena in order to predict probabilities. In general, they seem to do a good job (Lichtenstein, Fischhoff, and Phillips, 1977; Winkler and Murphy, 1973). In one experiment, their scoring rule values were often better than that of a Bayesian solution, because they understood some of the dependencies among the data whereas the Bayesian solution wrongly assumed conditional independence (Equation 3). Some forecasters seemed to "hedge" their results. Hedging involves using subjective scoring criteria (improper ones) which weight some mistakes as more disastrous than others (e.g., a farmer who does not prepare for frost is worse off than one who prepared for frost which did not happen) (Murphy and Winkler, 1971).

Forecasters were both accurate and consistent when they constructed credible intervals of maximum and minimum temperatures (i.e., the maximum would fall between x_1 and x_2 50% of the time). However, when they were told the interval and told to assess the probability, their estimates were excessive. In another experiment, they assessed area probabilities (e.g., probability of rain in San Francisco) and point probabilities (e.g., probability it will rain at station M within San Francisco). Their results indicated some statistical oddities. In some cases, point probabilities were larger than area probabilities $[P(A \cup B) < P(B)]$. Also, some forecasters use different definitions of events, making it difficult to interpret probability estimates because their meanings are variable (Murphy and Winkler, 1975).

PEATMOS is a computer aid which aggregates information and makes probability estimates for weather forecasting. Murphy and Winkler (1975) found that forecasters were not influenced by knowing the prediction made by this computer aid. Neither PEATMOS nor the forecasters showed a clear edge in estimation performance.

Forecasting is an ideal estimation situation since forecasters are not under great time pressure and they are professionals whose daily job involves probability estimation. Thus, they are not as likely as other estimators to resort to heuristics. However, even their performance shows some evidence of statistical biases.

Lichtenstein, Fischhoff, and Phillips (1977) surveyed human performance in a variety of different tasks and realistic situations, and concluded that the human operator was poorly calibrated. The operator was more likely to be excessive than conservative, but he showed both types of errors.

The human operator performs relatively well at tasks that involve "configural judgments" (Anderson, 1972). The operator is apparently able to see a pattern underlying a profusion of interrelated cues Medical diagnosis involves these types of judgments. However, generally computers are no worse than doctors at diagnosing diseases (Beach, 1975).

Staël von Holstein (1972) examined stock market predictions by stockbrokers, bankers, statisticians, etc., and found them to be overconfident (excessive) in their estimation. In fact, only 3 of the 72 subjects did better than would be expected by predictions based solely on the past frequency counts of the events being assessed. However, the consensus opinion (i.e., the weighted average of the best subject's predictions) did do better than predictions based on past frequency counts.

In a simulated military command situation, computer-aggregated $P(d_j|H_k)$ values were superior to operator aggregation. Here again, the thoroughly trained operator did relatively well compared to the computer. It was only when the incoming data were degraded that computer aggregation was clearly superior to the human operator (Schum, Goldstein, and Southard, 1966).

In summary, in some complex situations where time constraints are minimal and the operator has substantial knowledge of the area, his estimations can be characterized as fair to good (Beach, 1975; Winkler and Murphy, 1971). There also are a number of realistic tasks and concomitant situations (time constraints, stress, etc.) in which the operator can be characterized as almost a "non-estimator," i.e., his estimates not only are poorly calibrated, but also do not follow normative prescriptions (Kahneman and Tversky, 1972a,b; Lichtenstein, Fischhoff, and Slovic, 1977).

HELP IN ESTIMATING: AIDING AND TRAINING

Over twenty years ago, Simon introduced the concept of bounded rationality. Many formal decision models assume a rational decision-maker with unlimited processing capabilities. Simon argued that these models were unrealistic, since constraints on human computational, processing, and memory abilities made human decisions incompatible with normative predictions. However, if these constraints were understood and corrected for, then normative models could predict human performance. In other words, the human operator is rational within the bounds of these constraints.

In terms of the biases which the operator evinces in probability estimations, there are two possible ways to improve operator performance: (1) the operator can be trained, and (2) the operator can be aided.

Schaffer (1976) assessed the estimated probability distributions of subjects who had just finished a course in elementary statistics at the university level. The subjects were given feedback after each session and presumably understood the technical issues involved. However, training only improved their probability distribution estimations to a marginal extent. Also, the subjects' likelihood estimates (for Equation 2) became worse during training and their

posterior odds estimates remained at the same, rather poor level of performance. It seems that feedback which informed the subject that his distributions were excessive caused him to be too conservative in his likelihood estimates. The effect of feedback seems to be complex, making the estimation training of operators a difficult task.

However, Lichtenstein and Fischhoff (1978) trained subjects using a variety of feedback indexes including scoring rules, hit rate, and personal interviews, and found substantial improvement in their subjects' calibration. Since almost all the effect took place between the first and second session, this suggests that even a minimal amount of training can be helpful for at least some estimation tasks.

On the other hand, some of the biases seem particularly resistant to training (Fischhoff, 1975, 1977; Lyon and Slovic, 1976; Tversky and Kahneman, 1977) and in these cases aiding the operator is necessary. One of the first aids developed was the PIP aid mentioned previously. This aid computes posterior odds by aggregating likelihoods estimated by the operator. When there is a reasonable amount of independence among the estimations, this aid improves performance (Beach, 1975; Domas and Peterson, 1972; Schaffer, 1976). Although this aid extends the computational abilities of the operator, it leaves most of his biases uncorrected.

Recently, Slovic (1978) has reviewed a number of aiding techniques and general procedures for correcting biases. Kahneman and Tversky (1978a) introduced specific mathematical functions in order to correct a number of the biases they discovered in their earlier papers. Since these functions need few parametric inputs, they could be computerized and used as tactical aids (for situations where immediate decisions are unnecessary). More complex corrective functions have been suggested by Lindley, Tversky and Brown (1977). These functions are corrective procedures for probability distributions elicited from the operator which depart from probability axioms (e.g., probability of all possible events do not sum to one).

The three studies cited above represent a continuum from least to most sophisticated aiding procedures: (1) mostly insight (Slovic, 1978), (2) simple corrective functions (Kahneman and Tversky, 1978a), (3) complex functions (Lindley, Tversky and Brown, 1977). The usefulness of these procedures depends on both the particular situation and the training level of the operator. One past problem with decision aids is that operators, particularly those in leadership positions, are not apt to use them if they do not understand the output from the aid (Sinaiko, 1977). Thus simply understanding the biases (i.e., insight) is the necessary first step. Next, specific corrective procedures can be explained to the operator. If the operator understands the necessity and the mechanics of these procedures, they are more apt to use a preprogrammed aid which does the actual computation. Complex procedures such as those suggested by Lindley et al (1977) are probably more useful in situations where professional decision analysts are involved in the decision-making chain. Aiding and training are complementary processes, since the usefulness of an aid is some function of how well it relates to the training level of the operator.

SUMMARY AND OUTLINE

Table 1 is an attempt to outline the ODC in terms of various types of decision tasks operators perform when making probability estimates. The first two columns describe the characteristics of the paradigms used to study these tasks. Column 3 states the underlying model the operator should use for optimal solutions. Columns 4-7 characterize the operator's performance for these tasks. The most important sources for this characterization are given in Column 8.

"Event probability estimation (inexperienced)" refers to the operator's ability to estimate either probability distributions or odds that particular events will occur. It differs from the updating task because the operator may have little knowledge of either the probability structure of the events or the data which signal the events' occurrence. This task more closely approximates operator performance in situations in which he simply does not know the important parameters of the situation and yet must estimate the probability of an event occurring. For example, he might have to estimate the probability that route A is safer than route B the first time he flies in a hostile zone when there has been little advance intelligence.

On the other hand, "event probability estimation (expert)" refers to situations where the operator understands both the underlying dynamics of the situation and is somewhat familiar with the probability structure. The operator is assumed to be an expert at combining evidence with past frequency counts. The time stress is minimal. It differs from the updating task because the scenario is more realistic and the operator is well trained, thus understanding the possible dependencies which can exist among real world data. "Event probability estimations (expert)" are more likely to be made during mission planning than in mission execution. Obviously the two types of event probability estimations represent two extremes on a continuum rather than two discrete tasks.

TABLE 1. Summary of ODC for Decision Tasks Involved in Probability Estimation.

Decision task	Operator experience level	Face validity of experts	Normative model	Deviation characteristics	Heuristics	Biases	Performance level	References
Probability updating	Low	Poor	Equation 3	Conservative	LK = number of successes ÷ number of failures	1 misperception; 2 misaggregation; 3 response bias; 4 gambler's fallacy.	Poor to fair	Brickman & Pierce, 1972; Ducharme, 1970; Marks & Clarkson, 1972, 1973; Peterson & Beach, 1967.
Event probability estimation (inexperienced)	Low	Fair	Equations 1 and 2	Excessive	1 Availability; 2 representativeness; 3 anchoring; 4 concreteness	Ignore prior odds, set size, ...	Poor	Pitz, 1975; Slovic, Fischhoff, & Lichtenstein, 1977; Tversky & Kahneman, 1975.
Multistage inferences	Low	Good	Equations 7 and 8	Excessive	1 "As-if"; 2 "best guess", Equation 9	Ignore impact of less probable outcomes in first stage	Poor to fair	Gettys, Kelly, & Peterson, 1973; Schum & Pfeiffer, 1973; Snapper & Fryback, 1971.
Inferences based on nonindependent data	Low to high	Good	Equation 3 modified for dependencies	...	Unknown	...	Poor to good	Beach, 1975; Domas & Peterson, 1972; Zlotnich, 1968.
Probabilistic model generation	Low	Fair	Regression analysis	...	Linear trend assumption	Probability matching, gambler's fallacy	Poor	Brehmer, Kuylensteirng, & Liljergren, 1974; Brehmer, 1974; Lee, 1971.
Event probability estimation (expert)	High	Good	Subjective probability, Equation 1	Mixed	Unknown	...	Fair to good	Beach, 1975; Murphy & Winkler, 1975; Schum, Goldstein, & Southard, 1966; Winkler & Murphy, 1973.

WORTH ASSESSMENT

OVERVIEW

Besides assessing the probability of an event's occurrence, the operator must assess the worth of particular outcomes. This assessment should be independent of probability estimation. The first step in the assessment process is to generate a list of possible outcomes. Outcomes $[\phi_1 \ldots \phi_j \ldots \phi_n]$ are results of actions and are characterized by an n dimensional vector $[a_{11} \ldots a_{1k} \ldots a_{1n}]$ with each dimension representing a facet of the outcome's worth (Figure 5).

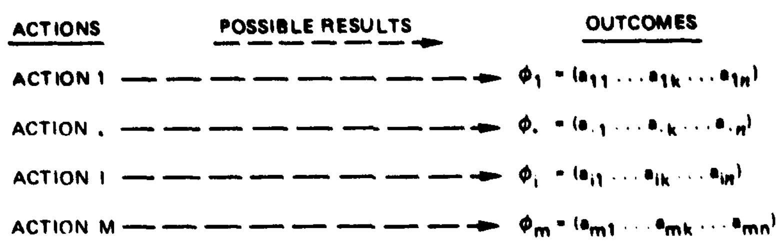


FIGURE 5. Structure of the Worth Assessment Task.

Formally, worth assessment is the evaluation of each outcome based on one or more dimensions of worth. The worth assessment process involves different scaling techniques which imply different measurement models. Worth assessment might be a complex, time-consuming process or a "snap judgment." The purpose of this section is to examine the formal properties of the process and to characterize operator performance.

The underlying assumption for all worth assessment is that the operator has a particular preference structure and a value for each outcome can be generated to represent it. Generating a value is far from trivial.

The pilot, in our example, must assess the consequences associated with sinking (or not sinking) the ship. Obviously, it is not an easy task. He must know what dimensions to consider and the ramifications of each outcome for these dimensions. These ramifications may have

complex second- or third-order reverberations (Slovic, 1979), so the pilot's evaluations may tend to be vague. Table 2 represents an example of possible dimensions and their evaluations for each outcome. However, as presently scaled these assessments are too ill-defined to base decisions on. Rather, they raise a number of questions. First of all, how valid are these values? What is moderate compared to high danger? Is there a trade-off between danger and military advantage?

TABLE 2. Worth Values for Sinking or Not Sinking Ship Imaged on Display.

Worth dimensions	Outcomes			
	Enemy ship		Friendly ship	
	Sink	Not sink	Sink	Not sink
A. Danger level to pilot	High	Moderate	None	None
B. General military situation	Improved	Same	Worsen	None
C. Damage cost to ship	$1M	Ø	-$2M	Ø
D. Change in fuel level	-30 gal	Ø	-30 gal	Ø

The formal procedures designed to answer these and related issues are reviewed in this section. Generally, these procedures consist of taking both the verbal and numerical values in Table 2 and converting them to a common numerical scale which reflects their relative worth. The actual value assigned to an outcome then depends on the measurement model the worth assessor feels is appropriate to the situation. In our particular situation, this process may seem too time-consuming to be of much value.

However, even in a time-constrained environment, the operator may informally consider different dimensions of the worth assessment situation and mentally make trade-offs among dimensions. Also, with the increasingly sophisticated software capabilities in attack aircraft, the actual worth functions could be programmed before the mission. The pilot could then enter values such as those in Table 2 assessing the immediate tactical situation and the computer could tabulate outcome values using the predetermined functions the pilot has generated before the mission. How valid such a procedure would be depends on issues discussed in the next section.

The next section will be devoted to techniques for generating worth functions. Then, measurement models these techniques are based on will be discussed. Finally, the operator's worth assessment characteristics will be examined.

SCALING TECHNIQUES

Methods

Scaling is the process of assigning numbers to items (Coombs, Dawes, and Tversky, 1971). Kneppreth et al (1974) have prepared an extensive, nontechnical guide to scaling of worth. A summary of their data is shown in Table 3. Techniques for scaling are listed below to give some of the flavor of what the worth assessment process consists of.

Ranking Method: In this method ranks are assigned to outcomes or an indication is made as to which of two outcomes is preferred. This gives ordinal scale information.

Equivalence Method: Outcomes are rated as fixed categories such as good, very good, etc. This technique also gives only ordinal information.

TABLE 3. Comparison of Worth Assessment Methods.[a]

Characteristics	Techniques: Ranking methods	Equivalence grouping	Direct methods	Gamble methods	Indifference methods
Probabilities required	No	No	No	Yes	No
Response required	Preference judgments	Indifference judgments	Qualitative judgment	Indifference judgments usually, but may be a preference judgment	Indifference judgments
Number of factor levels considered	Two	One	Two/Three	Three/Four	Two
Continuous or discrete factors	Discrete only	Discrete only	Either	Either	Continuous
Output	Relative worth	Approximate numerical worth	Numerical worth	Numerical worth	Ranking
Ease of use	Very simple	Very simple	Hard to do for some people	Tedious	Tedious
Training of decision maker	Almost none	Almost none	Moderate	Extensive	Little
Reaction of decision makers	High acceptability	High acceptability	Often requires selling for acceptability	Sometimes frustrating	Can be slow and taxing
Speed	Very quick	Very quick	Moderate	Slow	Slow
Data processing	Very little	Very little	Usually none	Moderate	Can be extensive
Unique worth assignment	No	No	Yes	Yes	Yes
Accuracy (Consistency)	High, if only a few alternatives	High, if only a few alternatives	Moderate	Accuracy is a function of patience, range of indifference, and understanding of probabilities and risk properties	Possible wide indifference area
Skill and experience required of analyst	Little	Little	Moderate	High	High

[a] From Kneppreth et al., 1974.

Direct Method: The levels of the dimensions are assigned numbers or graphed by the assessor against worth values. These techniques give interval information.

Gamble Method: Probabilities are varied between two outcomes until the assessor is indifferent. Alternately, worth levels are varied between two outcomes differing in probability until the assessor is indifferent. This technique gives interval information.

Indifference Methods: Indifference functions are graphed or elicited between levels of two dimensions. This technique can give interval information.

Keeney and Raiffa (1974) and Raiffa (1968) discuss gambling and indifference methods which are more sophisticated than the methods outlined by Kneppreth et al (1974). The main advantage to the more sophisticated methods is they can be used to generate worth functions of vectors with n dimensions rather than the few factors considered by Kneppreth et al (1974). However, these methods are based on the same basic elicitation procedures as the reviewed methods.

MacCrimmon (1968) reviewed a number of methods that consider the values in a multidimensional vector without exhaustively considering all dimensions for each outcome. These are listed below.

Lexicographic: Dimensions are ranked in importance. A subset of outcomes with high values on the most important dimension is chosen. This subset is further reduced by examining the next most important dimension and so on until the desired outcome is chosen.

Satisficing: A subset of dimensions is chosen and, if these values lead to a clear preference structure, the outcomes are ordered. However, if evaluation differences among outcomes do not surpass a criterial level, additional worth dimensions are examined.

Maximin and Minimax: Here the full set of dimensions are considered, but only the worst level among the dimensions is used to rank the desirability of the outcomes.

The relevance of these scaling techniques to operator performance depends on (1) practicality, (2) formal validity, and (3) psychological relevance.

Practicality and Scaling Techniques

One of the first practical problems facing the operator is whether to decompose an outcome into a multidimensional worth vector. In many cases, it may be too time-consuming and complicated to scale each

dimension rather than simply assigning a value to each outcome. However, in other cases, breaking down outcomes into their important dimensions helps the operator structure the dynamics of the problem in his own mind.

A practical solution to this problem is demonstrated with a worth assessment algorithm developed for group decision-making (Leal, Levin, Aamon, and Weltman, 1978; Leal and Pearl, 1977). The algorithm allows direct assessment of outcomes or decomposition, depending on the situation.

Outcomes are directly evaluated by each judge. A computer subroutine examines the range of the values, performing a sensitivity analysis. If the different judges' values lead to the same decision, no further analysis is done. On the other hand, if the divergence of opinion would lead to different decisions, each outcome is decomposed into a multidimensional worth vector. The values of the dimension are rated by each judge and discussed. This decomposition allows the decision-makers to identify and possibly mollify areas of discord. The amount of scaling done by the decision-maker is not pre-set, but rather it is the minimum necessary to arrive at a consensus. Sensitivity analysis is also possible for individual worth assessments, if the assessor gives ranges of worth and not a single value for each outcome.

MacCrimmon (1968) suggests combining a number of scaling techniques in order to simplify complex assessment problems. For example, a combination of lexicographic ordering and a minimax criterion could be used to eliminate obviously poor choices. The A-7E pilot might wish to consider 20 potential targets. He can choose the most important dimension (e.g., his safety) and attempt to minimize the maximum loss by eliminating any targets that have a good chance of shooting him down. For the remaining subset of targets, he can consider other important dimensions before making decisions. The important point is that these worth assessment techniques are not mutually exclusive: the best technique for a particular situation depends on the parameters of the situation and the ingenuity of the assessor.

Validity of Scaling Techniques

The validity of the scaling technique depends on (1) how the numbers were generated and (2) the reasonableness of the underlying measurement model. Considering the two-ship problem again should help in understanding these two concepts. It will be assumed that dimensions are additive. The assessment would consist of decomposing outcomes into a worth matrix (such as Table 2) and using one of the scaling methods to assign weights to various dimensions and numbers to various levels. Equation 9 could then be used to compute the worth of the Kth outcome (e.g., sinking ship B).

$$W_K = \sum_{i=1} b_i D_{ij} \tag{9}$$

where

W_K = value of the Kth outcome

b_i = weighting constant for dimension i

D_{ij} = value for the jth level of dimension i pertinent to the Kth outcome.

The validity of the worth assessment process depends both on the numbers generated and on the appropriateness of Equation 9. The first criterion (number generation) is a function of how well the numbers represent the operator's value system. Ranking both the dimensions and their weights results in losing interval information concerning the relative importance of different dimensions. Direct scaling techniques do not force the operator to make trade-offs. The assumption is that the operator can generate a number which represents a valid approximation of the dimension's relative worth. Indifference and gambling methods are more sophisticated. They generate interval information by forcing the operator to make indifference judgments between worth vectors by either changing values of dimensions or manipulating their respective probabilities. Those methods more nearly capture the essence of an additive model which is a trade-off function (Keeney and Raiffa, 1976; Raiffa, 1968).

However, the advantage of using the more sophisticated methods may be offset by their difficult and time-consuming nature. The choice of which technique to use depends on operator training level, time constraints, and desired precision, as well as on the theoretical validity of the technique (Kneppreth et al, 1974).

The second criterion involves Equation 9, which is based on an additive conjoint measurement model. The properties of the model must be realistic in terms of the empirical worth assessment situation. If the two-ship problem involved correlated dimensions or important interactions, an additive model could distort the true preference structure of the operator. The axiomatic basis of measurement models and their behavioral consequences are discussed in the next section.

Psychological Relevance

The ability of the operator to use these scaling methods and the processing load they impose on the operator determine their psychological relevance.

Simply ranking dimensions or outcomes involves less cognitive loading than does weighting dimensions or making paired comparisons. Such a simple strategy is justified for many decision problems, since empirical and analytical evidence indicates that more complicated methods often result in the same decisions (Eckenrode, 1965; McClelland, 1978).

There are a number of ways to generate numbers for direct scaling techniques (Galanter, 1965; Kneppreth et al, 1974). One of the most controversial is the ratio method. A low level is chosen arbitrarily, and other levels are assigned numbers in proportion to their respective values to the low level. Some active researchers favor using this method (Edwards, 1978), whereas Kneppreth et al (1974) argue that it is psychologically difficult for the operator to assess worth in this manner. Therefore, any scaling technique considered should be viewed with some caution. A technique should be chosen that is tailored to the individual operator by being simple and psychologically unequivocal to him.

A related concern is whether the very use of formal scaling models distorts the decision situation. It is possible that the values elicited during such a procedure are unstable and represent a conciliatory attempt by the operator to cooperate in the scaling process. This is particularly possible when the values are elicited from the operator by a decision analyist. The analyist's efforts to simplify the decision situation may result in structuring the problem in a manner that is totally unrepresentative of the situation as seen by the operator. One way to circumvent this is to have the operator generate an informal, intuitive analysis of the decision situation before any formal analysis is done. Then, if the operator's informal model differs from the formal scaling solution, the operator can rethink the problem and attempt to reconcile these discordant assessments. Using this strategy, the operator is not driven by the rigid formalism of some of the scaling techniques, but is able to use his own intuitive judgement in order to help guide the worth assessment process. In general, any scaling procedure that passively elicits values from the operator, runs the risk of dictating rather than discovering an underlying preference structure. (Fischoff, 1978; Fischoff, Slovic and Lichtenstein, in press.) Fischoff (in press) has recently compared worth assessment methods to clinical methods in psychology. The analogy is apt since, when both methods use insight and flexibility, the results are usually very useful. However, when the analyist is wedded to formal doctrine and inflexible models in either discipline, the resulting analysis often does more harm than good.

One purpose of decomposing outcomes into worth vectors is to reduce the operator's cognitive load. Assessing outcomes involves mentally integrating information over a number of dimensions and levels. Decomposition allows the operator to assess worth in smaller, more manageable chunks (Fischer, 1975; Miller, 1956).

MEASUREMENT MODELS

Importance

Measurement procedures (such as Equation 9) imply that the numbers being used have certain properties. The implicit properties underlying the measurement procedures are its axioms (Krantz, Luce, Suppes, and Tversky, 1971). A measurement model, then, is a collection of axioms that allow specified numeric procedures to take place among events defined by the model. Models are useful because they allow manipulation and abstraction of real-world processes. Pearl (1977) points out that worth assessment is related to specific neurological activities which govern the operator's preference structure. Measurement models are the framework in which these activities are assessed and assigned numbers.

The decision-making characteristics of the operator depend on how well the particular worth assessment model he uses reflects his preference structure. Part of the operator's task is circumscribed by the model's axioms. If his performance is incompatible with some of the axioms, then particular measurement procedures may be inappropriate.

Coombs, Dawes, and Tversky (1971) make the distinction between "simple" and "existential" axioms. The former deal with the model's specific assumptions, whereas the later involve conditions such as "continuity" dealing with properties of number systems in general. The present discussion will focus on "simple" axioms which have important behaviorial consequences; all necessary axioms for a particular model will not be discussed.

A thorough discussion of measurement models is beyond the scope of this paper. The focus of this section will be on (1) how realistic the assumptions of the model are in relation to operator performance and (2) the usefulness of procedures based on the model for the worth assessment tasks. For a complete discussion of the mathematical and logical foundations of these models, the reader is referred to Coombs, Dawes, and Tversky, 1971 and Krantz, Luce, Suppes, and Tversky, 1971.

Ordinal Preference Models

Ordinal preference models are the basis of scaling techniques that measure the operator's preference structure on an ordinal scale.

The first axiom, A.1 (connectivity), indicates that an ordinal number can be assigned to any two outcomes (x or y) indicating preference or indifference (Luce and Suppes, 1965).

$$x \succsim y \quad \text{or} \quad y \succsim x \tag{A.1}$$

This class of models also allows for the use of strict inequality operators (>,<) where clear preference structures are evident. A second axiom, A.2, requires that relations among outcomes be transitive. Transitivity means that if x is preferred to y and y to z then x must be preferred to z.

$$x > y \quad \text{and} \quad y > z \quad \text{then} \quad x > z \qquad \text{(A.2)}$$

Because of variability in human performance, A.1 and A.2 are certain to be violated in many real-life tasks. Tversky (1969) suggests that a weak stochastic transitivity condition, A.3, is the limiting factor for determining whether the assessment process is based on an ordinal preference model.

$$[x,y] > 1/2 \quad \text{and} \quad [x,z] > 1/2 \quad \text{then} \quad [x,z] > 1/2 \qquad \text{(A.3)}$$

where [x,y] means x is preferred to y.

Axiom A.3 indicates that, if the operator examined three outcomes a large number of times, his orderings would be transitive more than half the time. Most experimental results support models that are based on at least weak stochastic transitivity (Luce and Suppes, 1965; Tversky, 1969). However, for experiments with a large number of possible outcomes, only about 25%, at the most, of the relationships can be intransitive. Thus these tests may be insensitive to some patterns of intransitivity, because the patterns would show up in the data analysis very rarely. Tversky (1969) examined situations where he suspected intransitive relationships, and his results supported the hypothesis that the human operator is intransitive under certain conditions. Table 4 illustrates a condition similar to those in which Tversky found the operator to be systematically intransitive.

TABLE 4. Possible Intransitive Relations Among Ship-Destruction Preferences.

Ship	Worth dimensions	
	Enemy ship cost, $M	Danger level to pilot
A	9	Low
B	10	Moderate
C	12	High

The operator would choose to sink A rather than B, and B rather than C, but C rather than A. The rationale of the operator is that cost is the most important dimension, but that, unless the values in that dimension are at least $3M apart, the preference would be based on danger. In such a case no rank ordering of the outcomes is possible; thus any of the scaling techniques based on ordinal information are invalid.

Such an example is not contrived; there is other experimental evidence to support the notion that values must be a certain critical distance apart before clear preferences exist (Coombs, Dawes, and Tversky, 1971; Suppes and Walsh, 1959). Also, polling records show clear cases of systematic and not just nonstochastic intransitivity (Lee, 1971).

This does not indicate that the operator is necessarily irrational, only that worth assessment may involve heuristic rather than strictly analytical processes. The operator with a limited processing capacity may not be able to examine multidimensional worth vectors for each outcome. Thus he may lexicographically search dimension by dimension, eliminating outcomes with less than optimal values on important dimensions. Tversky (1972 a, b) provides data to support such a model of outcome preference, which he refers to as "Elimination by Aspect." This method may not be ideal, since important information in the multidimensional worth vector is ignored; however, it may be efficient considering the operator's processing limitations.

Ordinal preference models are the simplest measurement models for worth assessment. However, even for this relatively simple case, operator performance is not always commensurate with ordinal models. In certain circumstances (Luce and Suppes, 1965; Tversky, 1969), heuristic assessment procedures might be employed by the operator which violate transitivity assumptions. The main reason for this seems to be the operator's processing limitations rather than inherent irrationality.

Conjoint Measurement Models (Additive and Multiplicative)

Conjoint models are the basis of scaling techniques which measure the joint effect of two or more dimensions on the worth of an outcome. These models presuppose an ordinal model, but involve additional assumptions resulting in interval information concerning worth. Decomposition of an outcome into different additive dimensions is based on an additive conjoint model. The following two definitions represent the simplest additive model having two dimensions x and y, each having three levels: a, b, c and d, e, f, respectively (Coombs, Dawes, and Tversky, 1971).

$$\phi(ad) = h(a) + q(d) \tag{D-1}$$

$$\phi(ad) \gtrsim \phi(be) \quad \text{if} \quad M(ad) \gtrsim M(be) \tag{D-2}$$

where

$(x_i y_i)$ = vector containing the ith levels of x and y

ϕ,h,q = functions defined on x and y

$M(x_i y_i)$ = measured worth of $(x_i y_i)$

The first definition states that a vector is decomposable into two additive functions representing levels of its important dimensions. The second definition indicates that the decomposable function has the same ordinal properties as the measured effect of the nondecomposed worth vector.

Table 5 illustrates these two definitions. The levels (a, b, c and a, e, f) of the two dimensions (x = ship cost; y = danger to pilot) are assigned numbers reflecting their relative worth. These values are listed as marginal scale rates in the table, whereas the cell values reflect their measured joint effect. These definitions are satisfied if (1) the cell values are some additive function of the scaled values of cost (x) and danger (y), and (2) the value assigned an outcome is commensurate with the operator rankings of these outcomes. These definitions imply two types of independence.

The first type of independence requires that the marginal scale values uniquely determine the measured conjoint effect. This means that the dimensions do not have an interactive effect when combined. The second type of independence implies that different levels of the dimensions are separately realizable. This refers to conditions where having a particular value on x does not determine the level of y. For example, this type of independence would be violated if a value of high danger precluded a value of low cost as a possible outcome. (Krantz, Luce, Suppes, and Tversky, 1971)

TABLE 5. Additive Conjoint Effects of Levels of Dimensions x and y.*

Level		Dimension y (danger level to pilot)		
		d	e	f
Dimension x (ship cost)	Marginal scale values	17	19	38
a	2	19	21	40
b	0	17	19	38
c	−4	13	15	34

* Scaled values are some weighted functions of the decomposed values generated during the worth assessment process (see Keeny and Raitta, 1976) and cell entries are a measure of the joint effect of x_i and y_i (i.e., $M(x_i\ y_i)$).

Other necessary properties of this model include solvability and cancellation. Solvability assures that there are sufficient levels of the dimensions to compute a solution. Cancellation, A.4, is an algebraic result of the additive combinations of certain levels and can be interpreted as a special case of transitivity.

$$\begin{aligned} &\text{If } M(ae) \geq M(bd) \text{ and} \\ &M(bf) \geq M(ce) \text{ then} \\ &M(af) \geq M(cd) \end{aligned} \tag{A.4}$$

This relationship is not obvious and results from the fact that, if additivity holds, then $M(ae) \geq M(bd)$ and $M(bf) \geq M(ce)$ can be decomposed into additive functions and rearranged into Equations 10 and 10'.

$$h(a) - q(d) \geq w(b) - g(e) \tag{10}$$

$$w(b) - g(e) \geq l(c) - r(f) \tag{10'}$$

This rearrangement indicates that because of transitivity the following is true:

$$h(a) - q(d) \geq l(c) - r(f) \tag{10a}$$

Rearranging this equation into its additive form (10b) implies that, if definitions 1 (D-1) and 2 (D-2) are true, the result of Axiom [illegible] (10c) necessarily follows.

$$h(a) + r(f) \geq l(c) + q(d) \tag{10b}$$

$$M(af) \geq M(cd) \tag{10c}$$

The fact that this relationship is not intuitive makes experimental tests all the more crucial. Coombs and Komorita's (1958) subjects evaluated the worth of outcomes differing on the dimension of probability and monetary payoff. Their subject's ordering of outcomes was commensurate with the cancellation axiom in 29 of 30 cases. Adams and Fagot (1959) found that most of their subjects (80%) rated job applicants in accordance with the cancellation axiom most of the time (96%). Whereas, Fischer's (1976) results were similar for ratings of job opportunities.

These studies suggest that cancellation is a reasonable condition for at least the simple paradigms reviewed. Because of subject variability, these axioms are only an approximate description of operator performance. However, they still have proved to be good predictors of operators' decisions in a number of experimental paradigms (Fischer, 1975, 1976).

There are some cases when the underlying relationships among worth dimensions are simply not additive. Important interactions (especially nonmonotonic ones) and the situation where worth becomes zero if one dimension is zero are examples of such cases (Huber, 1974; Winterfeldt and Fischer, 1975). It can be shown that, if the values of dimensions are preferentially independent (see Keeney and Raiffa, 1976), a multiplicative conjoint model can be used to evaluate worth (Keeney and Raiffa, 1976; Keeney and Sicherman, 1976). Furthermore, experimental evidence suggests that additive models may be too simple to model the operator's preference structure in many worth evaluation situations (Anderson, 1965, 1970; Anderson and Shanteau, 1977; Grasser and Anderson, 1974).

Fischer (1976) studied worth assessment in both uncertain and certain environments. In both cases about half of his subjects' worth evaluations were better described by multiplicative than by additive conjoint models. However, only in the uncertain situation did the amount of variance accounted for appreciably change by using a multiplicative model.

In the final analysis, there is a trade-off between the simplicity of additive models and the added predictivity of using a multiplicative model. Many times, additive models prove to be adequate approximations of more complex processes simply because both models would lead to the same decision (Edwards, 1978; Gardiner and Edwards, 1975).

This discussion has assumed a worth vector with two dimensions. For cases when the number of dimensions (n) is greater than two, the property of cancellation is no longer necessary (Krantz, Luce, Suppes, and Tversky, 1971). Also, although there are 2^n-n-1 possible dimension combinations, Keeney and Raiffa (1976) demonstrate methods of selecting dimensions which are mutually independent by examining only a small subset of the possible combinations. Thus, in some ways, conjoint models become less complex for vectors with more than two dimensions.

Statistical Models

Another way to model the worth assessment process is to use statistical models based on the operator's actual preferences. In this case, the worth value for an outcome is generated by the operator and the weights for the dimensions in the worth vector are least-squares fits derived from the data.

Regression models were used to capture the judges' worth policies in early research on judgmental processes (Slovic and Lichtenstein, 1971). However, recent researchers have pointed out that this is an inappropriate use of correlational procedures (Anderson, 1972; Anderson and Shanteau, 1977; Slovic and Lichtenstein, 1971). Regression models are predictive rather functional models (Drapper and Smith, 1966). The underlying preference structure may be functionally complex, but still show a high correlation with a simple linear model. Anderson and Shanteau (1977) surveyed a number of areas of psychological research and concluded that using regression models to study functional processes leads to serious error even when the model was highly correlated with the criterion value.

Fischer (1976) studied worth assessment in both uncertain and certain environments. In both cases about half of his subjects' worth evaluations were better described by multiplicative than by additive conjoint models. However, only in the uncertain situation did the amount of variance accounted for appreciably change by using a multiplicative model.

In the final analysis, there is a trade-off between the simplicity of additive models and the added predictivity of using a multiplicative model. Many times, additive models prove to be adequate approximations of more complex processes simply because both models would lead to the same decision (Edwards, 1978; Gardiner and Edwards, 1975).

This discussion has assumed a worth vector with two dimensions. For cases when the number of dimensions (n) is greater than two, the property of cancellation is no longer necessary (Krantz, Luce, Suppes, and Tversky, 1971). Also, although there are 2^n-n-1 possible dimension combinations, Keeney and Raiffa (1976) demonstrate methods of selecting dimensions which are mutually independent by examining only a small subset of the possible combinations. Thus, in some ways, conjoint models become less complex for vectors with more than two dimensions.

Statistical Models

Another way to model the worth assessment process is to use statistical models based on the operator's actual preferences. In this case, the worth value for an outcome is generated by the operator and the weights for the dimensions in the worth vector are least-squares fits derived from the data.

Regression models were used to capture the judges' worth policies in early research on judgmental processes (Slovic and Lichtenstein, 1971). However, recent researchers have pointed out that this is an inappropriate use of correlational procedures (Anderson, 1972; Anderson and Shanteau, 1977; Slovic and Lichtenstein, 1971). Regression models are predictive rather functional models (Drapper and Smith, 1966). The underlying preference structure may be functionally complex, but still show a high correlation with a simple linear model. Anderson and Shanteau (1977) surveyed a number of areas of psychological research and concluded that using regression models to study functional processes leads to serious error even when the model was highly correlated with the criterion value.

On the other hand, the use of linear regression models to predict worth judgments has proved to be very successful (Dawes and Corrigan, 1974; Goldberg, 1970). Bootstrapping is the use of a regression model to predict decision rules based on operator performance. In general, bootstrapping models perform better than the judges they model (Dawes and Corrigan, 1974) because the variability reflecting lack of fit due to a linear model is often less than the variability due to the judges' inconsistencies (Goldberg, 1970).

Hamner and Carter (1975) used four regression models based on their subjects' estimates to predict ideal production rules. These models were linear combinations of different information sources such as sales forecasts, current inventory, etc. All four models led to production rules better than those chosen by the subjects whose data were used to generate the rules. Thus the weights inferred from the subjects' performance reflected a more stable indication of their worth vector than did the decisions made by the subjects. Dawes and Corrigan (1974) reviewed experiments in which bootstrap models of judges were more efficient than the judges' decisions in almost all cases.

Perceptronics (1977) has recently developed a decision aid (ADDAM) which incorporates features of bootstrapping models. The weights of the worth vector change as a function of operator decisions and are updated using an online computer program. Thus the operator has a computer model of his worth vector which can be used to select optimal decisions. Preliminary experimental evidence indicates that this aid improves operator performance.

Even simpler than fitted regression models are models with unit weights for each dimension. Dawes and Corrigan (1974) give evidence that such models result in predictions that are almost as good as least-square weights, provided there are no significant nonmonotonic interactions. This would greatly simplify the operator's task, since generating an additive model such as Equation 9 would consist of simply normalizing the levels of each dimension and then summing.

However, McClelland (1978) provides analytical arguments which show that unit weight models are most likely to be seriously erroneous for nondominated outcomes. Nondominated outcomes are those for which the operator must determine trade-off functions among dimensions in order to choose the outcome with the largest worth. Since nondominated outcomes are the most important for worth assessment, unit models would seem to have limited usefulness except in situations where the weights of the various dimensions are unknown.

Regression approaches extend the usefulness of using a worth vector by allowing weights for various dimensions to be determined by operator performance. The ultimate usefulness of these models depends on how

well they predict worth assessment in novel situations. Unfortunately, most of the models were tested on the data used to generate the model (Hamner and Carter, 1975).

Other Measurement Models

A number of multidimensional measurement models have been developed that depend on the concept of psychological distance (Coombs, Dawes, and Tversky, 1971). The worth models posit ideal points in multi-dimensional psychological space. Different metrics are used to measure the distance of worth vectors to these ideal points. The closer the vector to the ideal point, the greater the worth. Although scales based on distance models have been used in many areas of psychological research (Coombs, Dawes, and Tversky), quite a few of those scales violated the axioms they were based on (Beals, Krantz, and Tversky, 1968).

There has not been too much use of distance models for worth assessment, probably because these techniques have more theoretical than practical appeal. Klahr (1969) measured different dimensions of college admissions standards (e.g., college boards, I.Q., etc.) using the concept of psychological n-dimensional space with ideal points. His results predicted selections comparable with other scaling techniques, although the simpler additive techniques were better at predicting judges' decisions. Klahr argued that this approach still should be considered, since the values for worth were based on subjective rather than nominal metrics. Coombs (1958) was able to explain certain preferential inconsistencies by testing a model based on psychological distance.

Another approach to worth assessment is functional measurement scaling. These methods involve monotonic transformation of nominal scales to subjective scales and then testing various worth models (multiplicative, additive, etc.) using an analysis of variance. The main advantage to these methods is their dependence on operator behavior rather than axioms positing ideal behavioral patterns (Anderson, 1970; Anderson and Shanteau, 1970). However, functional measurement is more of a means of scaling and testing various conjoint measurement models than a separate measurement model in its own right.

EXPERIMENTAL COMPARISONS OF OPERATOR AND TASK WORTH ASSESSMENT CHARACTERISTICS

It is difficult to compare the various measurement procedures empirically, because the criterion value is not objective. Also, better performance for one procedure in a particular experimental paradigm does not necessarily indicate better performance under different conditions. Even comparing procedures' predicted results to actual decisions assumes that the operator's decisions perfectly reflect his preference structure, an assumption that most of the data contradict.

However, with these rather strong reservations in mind, empirical studies do indicate how well the human operator uses these procedures under specified conditions. In general, empirical studies show that decomposition procedures correlate well with both outcome ordering and regression approaches to worth assessment (Fischer, 1975).

Hoepfl and Huber (1970) compared two additive procedures with subjective levels of the dimensions being derived by their subjects. The criterion value was a worth value assigned to outcomes. For the first procedure, subjects directly scaled the weights for each dimension, whereas for the second approach the weights were fitted using a least-squares regression fit. The average correlation coefficient for the first approach was 0.87 and for the second 0.91 (both were corrected for bias). The important finding was that both models were able to predict the criterion value fairly well. Since the regression approach is a mathematical minimization of the residual sum of squares, it is not surprising that it accounted for more of the variance than the scaling procedure. To compare the two procedures adequately, the regression model would have to have been compared to criterion values other than those used to fit the model.

A tendency to give more uniform weights for all dimensions was noted for the scaling procedure. This tendency has been noted in other research on decomposition procedures and suggests that operators may be conservative when scaling (Edwards, 1978; Fischer, 1975). It appears as if the operator does not wish to assign weights with too high a proportion of the total value to any one dimension when he decomposes an outcome, whereas his actual decisions reflect heavy weighting of some dimensions.

Huber and Daneshgar (1971) compared five worth assessment procedures and found some interesting interactions. Regression weights were more reliable than decomposed scaling weights for experienced subjects in predicting job ratings. However, the opposite was true for inexperienced subjects. This suggests that decomposing job satisfaction into different dimensions helps when the subject is inexperienced in terms of the outcome, whereas experienced subjects are able to judge outcomes (i.e., jobs) directly more easily.

Another interesting find was that, in terms of ratings, a multiplicative model was better than an additive model, but, in terms of the job actually taken, the additive model was a better predictor. It is difficult to evaluate this because employment depends on more than just wanting a job, but it does suggest that simpler models may be better predictors for situations other than the one used to generate the model.

Humphreys and Humphreys (1975) compared psychological distance models and factor analysis with two additive decomposition models (one

with unit weight and one with weights derived using a gamble-type procedure). The criterion value was how well subjects liked six films. There were six dimensions in the worth vector. Additive decomposition models were better predictors of film choice, with the scaled weights being better than unit weights.

Satisficing reduces the cognitive load of the operator by reducing the scope of the multidimensional worth vector. There is little experimental evidence to suggest that subjects use this strategy. Sheridan, Richards, and Slocum (1975) compared a satisficing approach with an approach considering all dimensions of job worth, for nursing students' job selections. Their subjects' data did not support a satisficing approach to job selection. Ölander (1975) had his subjects choose either stocks (monetary value) or books (subjective value), and found no evidence that satisficing procedures were used in the former case and only limited evidence in the latter case.

Another way to reduce cognitive load is to choose outcomes with certain dimensional values rather than trying to integrate (or decompose) all the information in the worth vector. This would involve a lexicographic approach which Tversky (1972 a and b) refers to as "Elimination by Aspect" (EBA). EBA has two important behavioral consequences: (1) an outcome's worth changes as a function of the value of important dimensions for other outcomes, and (2) operators' cognitive loading is reduced. Tversky (1972 a) found support for the first hypothesis, suggesting that his subjects used a lexicographic approach.

For the experimental situations reviewed, additive decomposition was a relatively efficient means of assessing worth. The regression approach is promising, but is rarely used to predict decisions in novel situations. The use of satisficing and multidimensional "distance" techniques received little support from this review. However, this may be due more to lack of empirical research in this area than actual shortcomings of these methods. Evidence, both logical and empirical (Tversky, 1972 a and b), suggests that the operator uses lexicographic methods of worth assessment in certain situations. The latter finding suggests that, as in probability estimation, worth assessment may involve heuristic rather than strictly analytical processing.

SUMMARY OF WORTH ASSESSMENT PROCEDURES

Table 6 extends the task characteristics outlined in Table 3 by Kneppreth et al. The first column lists task characteristics. The more important measurement procedures are listed across the top of the table. Items A and C are self-explanatory. Item B is a rather rough measure of the difficulty of eliciting worth values from the operator. It is given as a range, since the difficulty varies, depending on

TABLE 6. Summary of Characteristics of Different Worth Assessment Procedures.

Task characteristics	Measurement procedures						
	Outcome ordering	Additive decomposition	Multiplicative decomposition	Regression	Lexicographic ordering	Satisficing	Distance measures
A. Normative model	Ordinal preference model	Additive conjoint model	Multiplicative conjoint model	Least-squares; statistical models			Multidimensional measurement models
B. Ease of use:							
Easy	X			X	X	X	
Medium	X	X	X	X	X	X	X
Difficult	X	X	X				X
C. Scaling technique:							
Ranking	X	X	X			X	
Paired comparison	X	X	X			X	X
Equivalent categories	X					X	
Direct		X	X			X	
Gamble		X	X			X	
Indifference		X	X			X	
Least squares fit				X			
Lexicographic					X		
D. Task performance characteristics:							
Analytical		X	X				X
Holistic	X						
Information lost	X				X	X	
Little analysis required	X				X	X	
Moderate analysis required		X					
Sophisticated analysis required			X	X			X
Reduces processing load				X	X	X	
Interpretation difficulty			X				X
E. Comments	Requires operator to mentally integrate over worth vector.	Weights for unimportant dimensions overestimated.		A) Predicative rather than functional model. B) May be poor in predicting future trends.	Similar to elimination by aspect choice model.		
F. Source	Luce & Suppes, 1965; Tversky, 1967.	Coombs, Dawes, & Tversky, 1971; Krantz, Luce, Suppes, & Tversky, 1971.	Keeney & Sicherman, 1976; Winterfeldt & Fischer, 1975.	Anderson & Shanteau, 1977; Drapper & Smith, 1967.	Tversky, 1967; 1972 A and B.	MacCrimmon, 1968; Ölander, 1975.	Beals, Krantz, & Tversky, 1968; Klahr, 1969.

the scaling procedure used, the number of outcomes, etc. For example, additive decomposition may be easier than outcome ordering if there are a large number of outcomes and few dimensions, while the opposite is true for few outcomes and many dimensions.

If the characteristics listed in C and D are descriptive of a particular procedure, an X is entered in the appropriate cell. Since not all the task performance characteristics are self-explanatory, a brief explanation is necessary.

Analytical refers to the nature of the worth assessment procedure. An analytical procedure breaks down an outcome into either a geometric representation (distance measured) or a functional representation (additive or multiplicative) of its constituent worth dimensions. A holistic task means the value of an outcome is judged without analyzing its constituent parts. Some procedures such as lexicographic ordering and satisficing are compromises between these two strategies. Information lost refers to loss of worth information because the worth vector is not completely analyzed.

Three levels of analysis of the worth values elicited from the operator are required. This does not refer to the worth assessment itself, but rather to the analysis done of worth values after elicitation. For example, a regression approach requires the operator only to generate the criterion value, whereas rather sophisticated regression analysis must be done to model the operator's worth vector. Thus there are two issues concerning the difficulty of the worth assessment task: (1) how easy is it to elicit a worth value from the operator, and (2) how much additional analysis is required after obtaining this value to determine the worth function? Sophisticated analysis requires either an expert in scaling (or statistical) techniques and/or a computer program.

Reduction in processing load indicates that the operator's processing load is kept to a minimum for a particular assessment problem. For difficult problems with time constraints, such approaches should be considered; otherwise the worth assessment process might exceed the limited processing capacity of the operator. Interpretation difficulty means that the worth function is difficult to understand because the functional form of the underlying model is complex. Not widely used is self-explanatory.

Table 6 indicates that none of the measurement procedures is superior to the others. The parameters of the worth assessment problem (such as expertise of the assessor, number of outcomes, time constraint, desired level of precision, etc.) determine the usefulness of the various procedures for a particular problem.

ACTION SELECTION

OVERVIEW

Figure 6 is an overview of the action selection task. The operators' actions (responses R_1 or R_2) depend on probabilities $P(H_1/d_j \ldots P(H_m/d_j)$ and outcome worth vectors $(\phi_1 \ldots \phi_n)$. Thus, after processing incoming data in terms of probability estimation and worth evaluations for possible outcomes, this information must be integrated to choose an appropriate action. This section will focus on optimal selection rules based on utility theory.

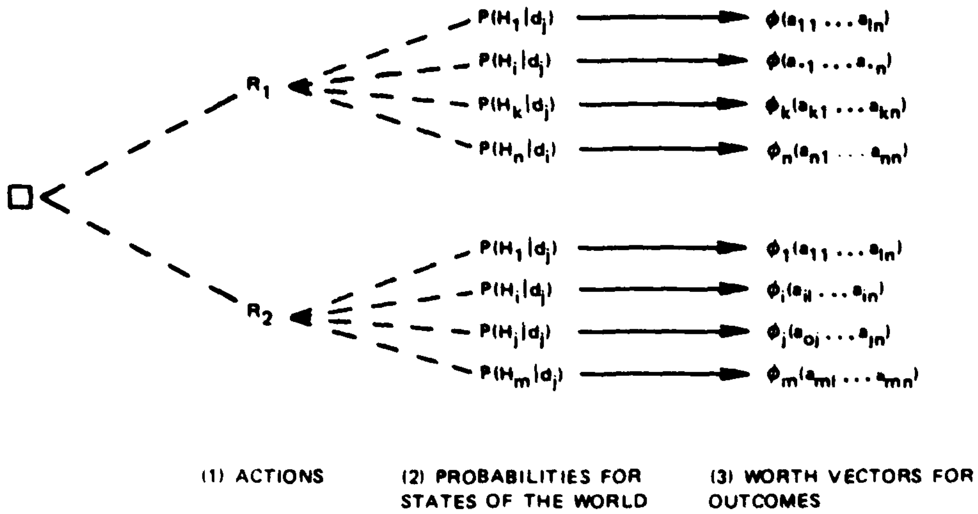


FIGURE 6. Decision Tree Showing Action Selection Task (R_1 or R_2).

Table 7 lists the probability and worth estimations the operator has generated.

TABLE 7. Utility and Uncertainty Information for Action Selection.

State of the real world.

Action	p Enemy	q Friendly
Sink	U(SE)	U(SF)
Not sink	U($\tilde{S}$E)	U($\tilde{S}$F)

The cell entries are utility (i.e., worth) functions measuring the worth vectors relating to the following outcomes: sink when enemy (SE), sink when friendly (SF), not sink enemy ($\tilde{S}$E) and not sink friendly ($\tilde{S}$F). Probabilities (p or q) of real-world states are listed over the columns. The operator's task is to find rules for action selection, based on the type of information listed in Table 7, which optimizes some decision criterion.

This entails finding utility functions for outcomes when the operator is unsure of what outcomes will result from his actions. In one sense, this involves conjoint worth measurement with probability being one dimension of the worth vector. However, the development of utility theory presupposes an uncertain environment which, in itself, puts unique processing demands on the operator (Kahneman and Tversky, 1978; Slovic and Lichtenstein, 1968).

UTILITY THEORY

Axioms

Utility is a measure of the worth of an outcome in an uncertain environment. For simplicity's sake, it will be assumed that we are comparing two outcomes X and Y, with X having probability p and Y having probability q of occurring. Six axioms are necessary to define the measurement model underlying a utility scale (Coombs, Dawes and Tversky, 1971). They will be discussed very briefly with emphasis being given to those axioms which have significant behavioral consequences.

Utility theory involves bets of the form (X, p, Y) or (Y, q, X) where q = 1-p. This states that X will occur with probability p and Y with probability q (or 1-p). Expected utility is a measure of the worth of such a bet. For example, how much money are you willing to pay for rolling a die which pays off X with probability p and pays nothing Y with probability q? The action selection problem can be thought of as representing two bets. For action "sink," the bet is of the form (SE, p, SF), whereas for action "not sink" the bet is ($\tilde{S}$E, p, $\tilde{S}$F).

The development of utility theory in terms of bets most likely has its roots in the historical interest of probability theorist in gambling situations. A U will suffix utility axioms. The first two axioms (U1 and U2) require the elements of the bet to be possible and the preference structure of the bets to be in accordance with an ordinal preference model.

U3 requires that the decision-maker be indifferent ($\sim$) between a simple bet and a compound bet having the same expected outcome.

$$\text{If} \quad (X,p,Y) = B, \quad \text{then} \quad (B,q,Y) \sim (X,pq,Y) \tag{U3}$$

A compound bet is a bet where the payoff is another bet. The probability of the operator obtaining B (which itself is a bet) is q. If he obtains B, he has probability p of obtaining X, thus the value of the compound bet must equal a simple bet for which the operator has a pq probability of obtaining X.

For example, X is $50 and Y is nothing. The operator should not be concerned with the form of the bet; only its expected outcome. A bet that requires two successes in a row in order to obtain $50 should have the same utility as a bet which pays $50 for one success, if the probability of a single success is 0.50 in the first case and 0.25 in the second case.

U4 indicates that the utility of two equally valued outcomes (X and Y) should not be affected if they are both involved in bets with another alternative.

$$\text{If} \quad X \sim Y, \quad \text{then} \quad (X,p,Z) \sim (Y,p,Z) \tag{U4}$$

U5 states that, if X is preferred to Y, then the value of any bet involving them is less than X and greater than Y.

$$\text{If} \quad X > Y \quad \text{then} \quad X > (X,p,Y) > Y \tag{U5}$$

U6 insures that an outcome can be made to be indifferent to a bet that involves a more preferred (X) and a less preferred outcome (Z) as alternatives.

If $X > Y > Z$, a p can be found for which

$$Y \sim (X,p,Z) \tag{U6}$$

The utility function for the various outcomes can be derived from U6. If X and Z are chosen so that X is the best outcome and Z is the worst outcome, then the p values elicited for intermediate outcomes would constitute a utility scale with X having a value of one and Z a value of zero (cf., Keeney and Raiffa, 1976; Raiffa, 1968).

In the terminology of gambling, actions with uncertain outcomes can be thought of as a bet in the form (X, p, Y). A utility scale allows the decision-maker to compute the expected utility of such a bet (Equation 11).

$$E(R_i) = p\ U(X) + q\ U(Y) \tag{11}$$

where U(X) and U(Y) are utility functions defined on X and Y.

The optimal decision-maker chooses actions which maximize expected utility. Thus the worth of actions in the problem presented in Table 7 can be expressed as:

$$E(S) = p\ U(SE) + q\ U(SF)$$

$$E(NS) = p\ U(\tilde{S}E) + q\ U(\tilde{S}F)$$

and the decision rule would be to choose the action with the higher expected utility.

Other Decision Rules Based on Utility Theory

In many real-life tasks, the odds (or probabilities) concerning possible states of the world change, but the utility values for different outcomes remain fairly stable. In such an environment the operator may wish to use a decision criterion based on odds. If the odds are above a certain level action, A is selected (R_A); whereas odds below the criterion favor the selection of action B(R_B). Such a criterion is referred to as β (Green and Swets, 1966). The advantage of β is that the operator can precompute (or computerize) utility functions and make decisions based only on his subjective probabilities (odds) concerning the changing state of the world.

To derive such a rule, odds must be found for which the optimal operator would be indifferent between R_a or R_b. This would be true when the expected utilities of the actions are equal.

$$E(R_a) = E(R_b)$$

or

$$p\ U(X) + q\ U(Y) = p\ U(Z) + q\ U(W)$$

or

$$p/q = (U(W) - U(Y))/(U(X) - U(Z))$$

or

$$\text{indifference } \underset{a-b}{\Omega} \text{ or } \beta = (U(W) - U(Y))/(U(X) - U(Z)) \qquad (12)$$

Equation 12 indicates that such a criterion can be developed, if the operator can generate the utility functions for possible outcomes.

Multi-Attribute Utility Theory (MAUT)

MAUT involves a conjoint worth measurement in an uncertain environment. The utility of an outcome is a function of the decomposed utility vector with each dimension of utility being called an attribute. A scaling technique based on U6 can be used to scale the multi-attribute worth vector (see Kenney and Raiffa, 1976; Raiffa, 1971).

Like other conjoint models, the worth function (in this case a utility function) is either a multiplicative or additive result of the values for each of the attributes. The resultant function must be commensurate with the utility axioms, and the utility function for separate attributes must be independent of the levels of other attributes (Keeney and Raiffa, 1976).

Different Criteria for Action Selection

Besides the maximum utility principle, there are a number of other criteria on which action selection can be based (Coombs, Dawes and Tversky, 1971).

Worth Selection: Only one outcome is considered possible for each action; thus only the worth of an outcome is considered.

Minimax: Outcomes assessed in terms of possible losses. The chosen action is that which results in the least undesirable outcome.

Maximin: Operator chooses the action for which the worst possible outcome has the highest value.

Maximax: Operator chooses the action for which the best outcome has the highest value.

Equal Probability: Operator has insufficient reason to assign probabilities to outcomes; therefore, he assumes all outcomes have the same probability.

All these methods can be criticized, because the long-run effect of using them would be a loss of utility (Lee, 1971). However, they all have the advantage of being relatively easy to use, since the operator does not have to estimate the probability. Some of these techniques may be especially useful for screening actions with undesirable outcomes.

OPERATOR PERFORMANCE

Introduction

Action selection is the use of both probability and worth information to select appropriate responses in terms of some decision criteria. The expectancy principle (expected utility) is the optimal solution. However, utility theory puts stringent constraints on the operator. The first issue addressed will be how well the operator's performance conforms to these constraints. Next, operator performance will be compared to other formal models of action selection. Finally, models based on the operator's information-processing abilities will be reviewed.

Performance in Terms of Utility Axioms

Because it is easy to think of utility axioms in terms of bets, most of the classical research on utility used a gambling paradigm to test its axioms. An obvious inconsistency between operator performance and utility theory involves the operator use of subjective probabilities; whereas the expected utility principle (EU) assumes that the probabilities are objective. Edwards (1955) and Savage (1954) proposed that utility theory be redefined to assume that utility axioms are based on subjective probabilities. Action selection, then, is based on the operator's attempt to maximize subjective expected utility (SEU) (Edwards, 1955). However, a perspective theory of operator performance based on SEU is at best a doubtful proposition.

One of the tenets of utility theory is path independence. This basically means that the operator is not concerned with the individual terms in Equation 11, only the result. This is partly a consequence of axiom U3. Experimental research indicates that compound bets are evaluated differently than simple bets directly contradicting U3 (Slovic, 1964). However, this is probably due more to the operator misaggregating probabilities than to any basic incompatibility of operator performance

and utility axioms. A more problematical finding for the concept of path independence is based on research done by Slovic and Tversky (1974). There are classical cases (developed by Allais and Ellsberg) where the operator chooses actions (bets) that are incompatible with the SEU principle, but are justifiable on rational grounds. One such example is illustrated in Table 8. In Slovic and Tversky's experiment, subjects were told to choose from gamble 1 or 2 in situation A and gamble 3 or 4 in situation B. A majority of subjects chose gamble 1 and gamble 4. However, no matter what subjective probability and utility function a subject uses, these choices are incompatible with SEU (Raiffa, 1971). This can be demonstrated by assigning values to end-points for these two scales; for 1.00 on the probability scale use 1.00 and for no money on the utility scale use zero. The symbols in parentheses can then be used to generalize the results to any utility or subjective probability function. According to SEU, preferring gamble 1 is represented by the following inequality.

$$a > pe + qa$$

or

$$a(1-q) > pe$$

However, preferring gamble 4 to gamble 3 implies the following inequality.

$$a(1-q) < pe$$

TABLE 8. Probabilities and Payoffs for Allais' Gambling Situation With Utilities and Subjective Probabilities in Parentheses.

Gamble	Probability	(pi)	Payoff, dollars	(U)
		Situation A		
1	1.00	(1)	1 million	(a)
2	.10	(p)	5 million	(e)
	.89	(q)	1 million	(a)
	.01	(r)	Nothing	(0)
		Situation B		
3	.11	(1-q)	1 million	(a)
	.89	(q)	Nothing	(0)
4	.10	(p)	5 million	(e)
	.90	(1-p)	Nothing	(0)

Since these preference structures are incompatible, most subjects did not choose according to SEU. What makes this paradigm important is that, even after explaining the reasoning behind the SEU interpretations, most of Slovic and Tversky's subjects still preferred 1 to 2 and 4 to 3. Thus, it can be argued that it is not the operator's lack of information that is responsible for his choices, but the fact that he has different criteria than SEU for action selection. Kahneman and Tversky (1978) use similar examples to further demonstrate that the operator's preference structure is not necessarily consistent with SEU.

Axiom U6 can be interpreted as stating that every bet has a certainty equivalent (or cash equivalent). Thus for every bet of the form (X,p,Y) there is some utility value (or perhaps money) which the operator would consider equal to the bet. Becker, Groot and Marschak (1964) manipulated the form of the bet in order to have a number of bets with the same cash equivalent. The variation among their subject's cash equivalents, in situations for which they should have been equal, demonstrated quite nicely that the relationship between an uncertain situation and a certainty equivalent is not apparent to most operators.

Axiom U4 implies that the context (i.e., values of other outcomes) of the bet should not change the utility value of outcomes. However, experimental evidence indicates that subjects consider the context of the bet and not simply isolated utility functions when making choices (Coombs, Dawes, and Tversky, 1971; Tversky, 1972a).

Equation 11 implies that the operator makes independent estimates of probability and utility in order to compute SEU (Savage, 1954). However, Slovic (1966) gives evidence that the utility of an outcome influences the operator's subjective probabilities. Some subjects make optimistic probability estimates in light of a big payoff, whereas others give pessimistic estimates. Thus not only do operators' performance characteristics run contrary to SEU, but also these characteristics appear dependent on personality differences.

In general, operator performance differs from utility axioms in many important aspects. Still it is possible that action selection based on SEU, although theoretically deficient, is a good predictor of operator performance. The next section will compare SEU to other possible worth assessment models.

SEU as a Measurement Model

SEU theory states that operators attempt to maximize average (or long run) worth (or utility) by using a selection model which weights worth estimates with probability estimates. However, the operator may view the selection task in terms of different processes than those posited by utility theory. The simple bet paradigm can be represented

by a vector with four terms (X,p,Y,q). SEU is one model of integrating these terms to assess the worth of different bets (i.e., actions). However, empirical results indicate that the operator considers more features of the action selection problem than implied by the SEU model and integrates the dimensions differently.

The variance, skewness, "gambling" situation itself and type of response all influence the worth assessments of different actions (Coombs and Pruitt, 1955; Lee, 1971; Royden and Walsh, 1959; Slovic, 1972; and Tversky, 1967). Thus the operator considers more than just possible outcomes and their probabilities when choosing actions. Furthermore, even the manner in which the operator uses the worth vector to assess the value of particular actions differs from SEU theory. Different operators focus on different dimensions of the worth vector, depending on both personality differences and type of response (Slovic and Lichtenstein, 1968). Apparently, the operator's initial determination of worth is based on the dimension (e.g., payoff, probability of loss, etc.) he feels is most important, Next, the operator uses some method to integrate the other information in the worth vector with the most important dimension in order to make a final evaluation (Slovic and Lichtenstein, 1968). This suggests that no formal measurement model could accurately describe the operator's action selection process unless assumptions are made concerning the operator's information-processing capabilities. Furthermore, the operator is limited in his ability to resolve utility differences among outcomes (Luce and Shipley, 1962; Suppes and Walsh, 1959).

It is not clear exactly what process the operator uses to integrate the worth vector. The averaging process implied by SEU theory predicts operator performance in at least an approximate sense (Anderson and Shanteau, 1970; Tversky, 1967). However, the resulting value is subadditive (i.e., $pX + qY < SEU$) (Shanteau, 1974).

In summary, a model that explains operator action selection must consider the following: (1) other factors besides outcomes and probabilities; (2) processing limitations of the operator; and (3) subadditive averaging process. Models which focus on these issues will be summarized in the next section.

<u>Models of Operator Action Selection Performance</u>

Models of human information processing identify three serial processing stages: stimulus encoding, cognitive processing, and response selection (Briggs and Swanson, 1969; Sternberg, 1969). Cognition involves the structuring of encoded stimuli into meaningful categories, whereas response selection involves translating these stimulus categories into the chosen response. There seem to be two substages of cognition involved in decision-making. The first stage involves a general

categorization and evaluation of possible actions, whereas the second stage is concerned with the formation of decision rules which result in response selection (Slovic and Lichtenstein, 1968).

Kahneman and Tversky (1978) propose a model of decision-making they refer to as "prospect" theory. A "prospect" is the worth and probability values associated with an action. After encoding these values, the operator categorizes prospects using various processing strategies such as recoding, combining, and segregating different dimensions of the prospect. Next, an empirical function is used to weigh probabilities ($\pi(p)$) with small probabilities being "over-valued," but most of the probability range is "under-valued" compared to riskless prospects. This process helps explain Allais' example. Since 1.00 would be over-valued in terms of the probability scale, such a weighting function could result in the observed preferences of Slovic and Tversky's subjects. Also, such a function would explain subadditivity.

The decision rule the operator uses for action selection is similar to the expectancy principle, but it focuses on the operator's potential change in fortune. In simple bet cases, when $X > Y > 0$ or $X < Y < 0$, Equation 13 is used to determine the operator's evaluation of prospects.

$$Vk = Y + \pi(p)\ (X - Y) \tag{13}$$

where Y is the minimum loss or gain and Vk is the worth of the kth action (or prospect).

Y is the amount the operator gains or loses for certain, with the second term of Equation 13 measuring the operator's potential change in fortune depending on the result of the action. This model of decision-making is used to explain empirical functions rather than to posit ideal functions. It explains most of the observed discrepancies between operator performance and SEU theory and serves as a useful approximation of operator performance. It is not as elegant as SEU theory, since it introduces both additional parameters ($\pi(p)$) and additional assumptions concerning operator processing limitations, in order to explain empirical results. Also, probability estimation and worth evaluation are not specifically addressed by the model. However, the advantage of such a model is that it attempts to describe decision-making in terms of operator performance characteristics instead of ideal performance criteria which may have little empirical validity.

Risk-averse functions are a good example of the difference between a normative and empirical approach to model building. Most operators show a tendency towards being risk-averse. That is, they would pay more for a sure thing than for an equivalent bet with the same expected value. Pratt (1964) and Keeney and Raiffa (1976) introduce a number of utility

functions to describe risk-averse behavior which are commensurate with utility theory. However, Tversky and Kahneman (1978) point out that most empirically derived risk functions are not commensurate with utility theory. Their approach was to fit the model to the data rather than vice versa.

Another model that has both normative and descriptive appeal is the "EBA" model mentioned previously. The probability of choosing an action is proportional to the lexicographic weights of different dimensions in the worth vector. Empirically, the model is supported by the tendency of the operator to focus on important dimensions rather than to compute trade-off functions (Slovic and Lichtenstein, 1968; Tversky, 1972 a and b).

SUMMARY

The maximum expected utility principle was shown to be the normative criterion for action selection. Using this criterion, decision rules were derived for outcomes, odds and multi-attribute decision problems. Subjective expected utility (SEU) was examined as a possible descriptive model of operator performance. Both because of serious violations of axioms and processing limitations of the operator, empirically derived models such as "prospect" theory were found to be superior to a model based on SEU theory.

Table 9 summarizes the performance aspects of different action selection models.

TABLE 9. Performance Characteristics for Action Selection.

Action selection model	Normative model	Maximize average value	Type of worth function	Type of probability function	Type of multi-dimensional evaluation	Lighten processing load of operator	Deviations of operative performance	Drawbacks
SEU	Utility theory	Yes	Utility function	Subjective probability	MAUT	No	(1) Axioms, especially U3, U4, and U6 are unrealistic. (2) Hypothesized worth vector incomplete. (3) Subadditivity. (4) Empirical risk function incompatible with theoretical function.	(1) Heavy processing load on operator. (2) Empirical performance not compatible with model.
Minimax Maximin Maximax Equal probability Worth selection	Game theory	No	Undefined	None	Undefined	Yes	...	(1) Decision rules do not maximize worth. (2) Ignore important worth and probability information.
Prospect theory	Assumes operator maximizes change in fortune	May or may not maximize average value	Utility function	(1) Subjective probability (2) Weighting functions	MAUT	Yes	...	(1) Model makes many assumptions and posits free parameters. (2) It has not been tested in situations other than those used to generate the model.
EBA	Lexicographic search	No	Undefined	Subjective probability	Undefined	Yes	Operator does use tradeoff functions in many situations.	(1) Empirical validity limited to specific situations. (2) Decision rules do not maximize worth.

INFORMATION SELECTION

OVERVIEW

The iterative nature of information selection can be seen by examining the feedback loop in Figure 2. Before selecting an action, the operator has the option of sampling more data, depending on both his own state of uncertainty and the cost of sampling. Cost is used in a general sense indicating cost in terms of physical and temporal constraints as well as monetary value.

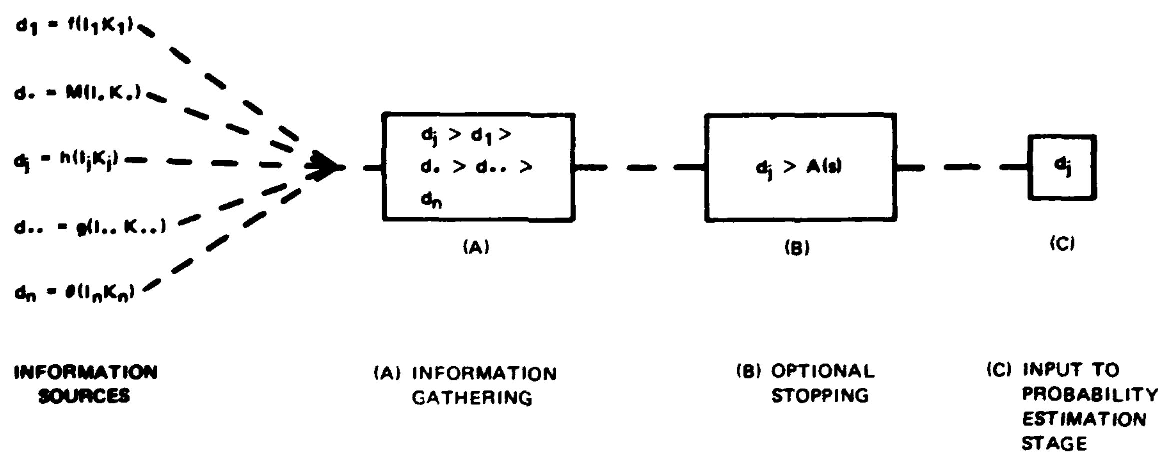
The internal dynamics of information selection are illustrated in Figure 7. Information selection involves decisions concerning both "how much" and "what type" of information to select (Nickerson and Feehrer, 1974). The latter process is designated as information gathering (Box A in Figure 7). This involves a selection among various information sources as a function of average information ($\bar{I}_j$) and cost of sampling (C_j) for each source. Formally, information is a measure of the reduction in operator uncertainty after sampling a datum (Equation 14) and average information is the expected informational value of data from a particular source before sampling.

$$I_j = -\log p(d_j | H_k) \tag{14}$$

where H_k is the hypothesis reflecting the actual state of the world.

In terms of our example, information gathering would consist of deciding which sensor devices to use and whether to get an updated intelligence report from an E-2 (tracker aircraft) or strike command. Information in its technical sense would be measured in terms of the increase in probability for the true hypothesis that each information source would result in. Cost would depend on such factors as time to retrieve information, danger involved in breaking radio silence, etc.

Since Equation 14 is simply a transformation of probability, information can be interpreted in terms of Bayesian statistics (Analytics, 1976; Savage, 1954). For example, diagnosticity, in Bayesian terms, is simply the algebraic difference between the informational value of the denominator and the informational value of the numerator of the likelihood ratio. There are a number of advantages in using information metric to measure operator performance. The metric is independent of the type of data selected and experimental paradigm used. Also, this metric has been used to measure human performance for a wide variety of cognitive and motor tasks, allowing operator performance in information selection to be compared to other human performance results (Briggs, 1974; Fitts, 1954; Hicks, 1951).



d_j = EXPECTED VALUE OF INFORMATION SOURCES
I_j = INFORMATION VALUE OF jTH SOURCE
K_j = COST PER UNIT OF INFORMATION FOR jTH SOURCE

FIGURE 7. Outline of Information Selection Task.

The second process involved in information selection is optional stopping (Box B). The "fair cost" of sampling a data source is the difference in value of selecting an action before and after sampling (Wendt, 1969). The optional stopping task involves deciding whether the "fair cost" of sampling the favored data source is more or less than the actual cost of sampling. If it is more, then sampling takes place, otherwise the operator stops sampling and selects the action with the highest expected utility

The problem our exemplary pilot would have concerns his trade-off between uncertainty reduction and danger to his aircraft. The closer he gets to the ship (and thus the more information he samples) the more likely the ship is to fire antiaircraft missiles at him. Also, of course, the more likely he would be to identify the ship. The optional stopping task consists of deciding when to stop gathering more information and either launch the missile or stop flying toward the ship and continue on his original flight plan.

Thus information selection involves both the first (A) and last (B) thing an operator does before committing himself to a course of action (i.e., (a) he must decide which data to sample, and (b) he must decide when to stop sampling and start acting).

EMPIRICAL RESULTS

Overview

Most of the experimental paradigms reviewed were similar to those used for the Bayesian updating tasks. The main difference consisted in the dependent variables which are number of data sampled and their informational value for the present task.

In general, subjects sample both too much and too little information depending on the experimental parameters (Nickerson and Feehrer, 1976). When the information value of the data source is too low, too much sampling takes place; whereas too little sampling takes place for information sources with relatively high values (Nickerson and Feehrer, 1976; Snapper and Peterson, 1971). There seem to be individual differences among operators, with some being overly cautious (want too much information) and others too incautious (sample too little information) (Edwards and Slovic, 1965). Also, there seems to be a nonoptimal relationship between the amount of information sampled and the amount available. Apparently, the operator is conservative. He does not want to deplete his information source so he weights his estimate of the "fair cost" of sampling as a function of the sample size available (Levine, Samet and Brahlek, 1975; Pitz, 1968).

The operator is able to use estimates of prior odds, diagnosticity and cost to determine when optional stopping should take place, but his actual criteria are not optimal (Wendt, 1969). The operator is able to learn how to be more efficient in an information gathering task, but performance is still less than optimal after practice (Levine and Samet, 1973). Optional stopping seems to be more time-consuming than information gathering (Levine and Samet, 1973).

An experiment by Pitz (1968) shows the relationship of operator performance for optional stopping and Bayesian updating. His subjects sampled a limited number of red and blue poker chips from two bags. This was similar in design to the Urn problem in that the probability of a sample being from a particular bag was a function of the difference (d*) in the number of red and blue chips (the Urn example is reviewed in the probability estimation section). If cost were held constant, the criterion for optional stopping should be based on d*. However, subjects used a criterion more sensitive to the proportion of red and blue data samples than to d*. This directly verifies the previously reviewed results of Marks and Clarkson (1973) which indicated that subjects estimated probability by proportion of different colored chips and not by the Bayesian solution based on d*.

The point of the comparison is that the results should be the same. The processing requirements for estimation of informational value and probability are identical in many respects. Therefore, when the other parameters of the situation are similar, we should expect the operators decision-making characteristics to be similar for these two tasks. This suggests that heuristics used for probability estimation are most likely evinced by the operator in information selection tasks also.

Time Constraints

It was mentioned previously that time constraints could be considered as a cost factor. Pachella and Pew (1968) give empirical evidence that the operator uses both time and monetary cost criteria for optional stopping in a choice reaction time task. One of the more interesting findings concerning the operator decision to stop sampling and start findings concerning the operator decision to stop sampling and start responding is the so-called "speed-accuracy" trade-off (Fitts, 1955; Pew 1969). This trade-off implies that the operator is able to sample information at a constant rate, but chooses different "informational" or "time" criteria depending on whether the task pays off for speed or accuracy. Apparently, when the operator is told to "speed up," he samples information at the same rate, but samples for a shorter time segment. The result of such a trade-off is that operator performance will be more error-prone in circumstances where he speeds up because he can only sample a given amount of information in a particular time segment. In the most extreme case, when he is under great time pressure, the operator may not sample information at all, but respond

randomly (i.e., guess) (Swensson and Edwards, 1971; Yelloit, 1971). If the "speed-accuracy operating characteristics" for particular operators are known, then error rate can be predicted for a task that has a temporal deadline attached to it (Pew, 1969).

Although the parameters of the "speed-accuracy" trade-off have been investigated using different choice reaction time paradigms, the real-world validity of such paradigms is suspect. Research needs to be done to determine the "speed-accuracy operating characteristics" for different operators using real-world scenarios when such factors as work load, stress, and payoff schedules are varied concomitantly.

SUMMARY: INFORMATION SELECTION

Information selection depends on criterion based on both cost evaluation (worth assessment) and uncertainty evaluation (probability estimation). Therefore, the operator's information selection characteristics depends on his decision-making characteristics for the other stages of decision-making. Heuristics and biases found in these stages are also pertinent to information selection performance.

However, there are a number of performance characteristics specific to information selection:

1. When there are a great number of information sources to choose from, operator performance is degraded. There seem to be two reasons for this: (a) the operator pays more than the fair cost for information, and (b) too much information overloads the operator's processing capacity.

2. Too sparse informational sources results in degraded performance for the opposite reason: (a) the operator will not pay for sources even when their cost is less than the fair cost, and (b) the operator tends to conserve sources rather than purchase needed information.

3. The operator can trade off the cost associated with time for information depending on the cost parameters for each.

ISSUES AND TENTATIVE CONCLUSIONS

OPERATOR AND TASK CHARACTERISTICS

This paper has reviewed operator and tasks performance characteristics for four stages of decision-making: information selection, probability estimation, worth assessment, and action selection. The most important general finding is that the operator is best characterized in terms of "bounded rationality" rather than in terms of normative or optimal models. It is necessary to understand the operator's processing limitations in order to understand how he performs as a decision-maker in diverse environments.

The concept of "heuristics" borrowed from the discipline of artificial intelligence (Simon, 1970) seems to be the most promising description of operator decision-making performance. A heuristic is an algorithm or rule which attempts to find satisfactory solutions without analytically examining all possible solutions. Heuristic techniques are used because of the processing limitations inherent in a particular decision problem. It is an apt description of operator performance for a variety of reasons:

1. Heuristic rules lead to solutions which are rational but not necessarily optimal.

2. The reason a heuristic is used has to do with the limitations of the processor.

3. Different heuristics are used for different tasks.

4. Heuristics lead to predictable errors.

The operator is constrained both by his processing limitations and his misunderstanding of normative processes. A number of biases and performance characteristics have been outlined in this paper, and rather than list these biases again, the following performance characteristics seem to best summarize the results:

1. Operators are not "conservative" or "excessive" estimators of probability. Rather, their particular performance characteristics depend on the parameters of the estimation tasks.

2. Operator biases are a result of adaptive rather than irrational processes. They are an attempt to make sense of a noisy environment.

3. Since some of these biases are predictable and fairly consistent, it is possible to use corrective functions that result in nearly optimal predictions.

4. Given sufficient time, the operator is able to decompose outcomes and develop trade-off functions which weigh the effects of a number of different dimensions in order to compute the worth for an outcome.

5. However, when the number of dimensions is large or time is constrained, the operator chooses assessment techniques commensurate with his processing limitations (e.g., lexicographic, satisficing...)

6. Action selection models which are based on empirical functions such as Prospect theory are better predictors of performance than those which posit mainly normative functions such as utility theory.

7. The operator performs best in information selection tasks when there is a moderate amount of information to select from. Too much or too little information results in degraded performance.

IMPROVING PERFORMANCE THROUGH TRAINING AND AIDING

The paper has identified a number of operator performance deficiencies. Training and aiding are two possible ways to improve operator performance. Techniques for improving probability estimation and worth evaluation (e.g., bootstrapping) performance have been discussed, this section will discuss more global approaches to performance enhancement in decision-making.

Saleh et al (1978) have developed a decision-making program imbedded in the anti-submarine warfare (ASW) context which trains the operator in all aspects of decision-making. The important decisions the operator must make in this environment are pinpointed and defined as generic decision tasks (e.g., action selection, probability estimation...). Then, the operator is trained to use decision analytic techniques to make decisions for these important decision points from the ASW environment. The premise of this experimental program is that the operator is being trained to use powerful decision analytic tools (decision trees, probability estimations...) while at the same time he is trained to solve specific decision problems in the ASW environment. Since this project is still in its early stages, it is too early to evaluate its success.

Crooks, Kuppin and Freedy (1977) have introduced the concept of decision-aiding into the training domain. They trained troubleshooters in fault-finding for electronic circuits. The aiding device gave feedback and/or aiding during the task. Feedback consisted of telling the subject what the expert would have done after the decision was made. Whereas, aiding involved telling the subject what alternatives the expert would have considered before the decision was made. A combination of aiding and feedback trained the operator in making less costly decisions even in situations where no aiding or feedback was available.

However, in general, training using decision analytic techniques or aids is still in the exploratory stages. The advantage to such training seems obvious, but it remains to be seen whether the results live up to its promise. In a similar vein, a number of decision aids are being developed to help the operator in operational settings, particularly in the command and control environment. Preliminary results indicate that devices which result in action selection are not as practical as devices which give the operator information to aid him in making the choice. The former devices tell rather than inform the operator, leaving him uncertain as to the logic of the chosen decision.

Two examples of the "information" type device are an Emission Control (ECOM) decision aid (Decision-Science, 1978) and a device to aid in selecting optimal air strike routes (Integrated Science, 1978). In both cases, the operator chooses different feasible alternatives, sensor configurations in one case and air strike routes in the other, and the aids display the important parameters for each choice and compute their expected utility using criteria the operator is familiar with. This allows the operator to explore different alternatives he has generated himself and allows the computer to use its computational power to enhance the operator's decision-making performance.

In summary, both aiding and training seem promising as a way to enhance the operator decision-making skills. The approaches are complementary, since aids can improve training and training is necessary for the operator if he is to appreciate the advantages of aids. Both approaches are still in the developmental stage. If they are to prove useful, their purpose will be to enhance rather than to replace the operator as a decision-maker.

RELATED ISSUES AND FUTURE TRENDS

Although hypothesis and outcome generation was discussed in this report, very little was said concerning their performance characteristics. This is not meant to suggest that these processes are unimportant, only that not much is known concerning operator performance for these processes. Preliminary evidence suggests that the operator considers

far too few hypotheses (Gettys, Fisher, and Mehle, 1978) and outcomes when structuring the decision problem. The problem of structuring the decision situation may prove to be the most important component of decision-making and should prove to be a fertile ground for future research (Saleh et al, 1978).

Most of the aiding and training programs mentioned have their impetus from a new applied science referred to as "Decision Analysis." Decision analysis "merges the logical foundations of statistical decision theory with the capabilities...developed in the fields of system analysis and operations research" (Matheson, 1969). Its main impact has been on the business community (Brown, Kahr, and Peterson, 1974), but the techniques developed by decision analysts are widely applicable and have been used in many military situations. The major drawback to decision analysis is that it demands a high level of expertise and therefore a formal decision analysis should be reserved for projects with a high potential payoff. However, the technology of aiding and training decision-makers will continue to benefit from the research results generated by this discipline.

Artificial intelligence (AI) is another discipline which has a potentially important impact on decision-making. AI involves problem solving algorithms (i.e., techniques) for situations where the solution demands complex logical decisions such as chess, geometric recognition, etc. In other words, its domain is problems previously thought to be solvable only through human intelligence.

One of the more applied uses of AI is knowledge engineering (Figenbaum, 1978; Shortliffe and Buchanan, 1975). Basically, this is the logical structuring of rules elicited from experts in a particular field to model the experts' performance in some decision task. Computer programs using knowledge based on algorithms have been successfully used for disease diagnosis, chemical analysis, and in military situations to predict the presence of possible hostile forces. Knowledge based algorithms are currently being used for aiding in only a few military situations. However, the potential applications for military problems are great, especially with the rapidly increasing technology of both computer hardware and software.

BIBLIOGRAPHY

Adams, E. and R. Fagot. "A Model of Riskless Choice," *Behav. Sci.*, Vol. 4 (1959), pp. 1-10.

Anderson, N. H. "Averaging Versus Adding as a Stimulus-Combination Rule in Impression Formation," *J. Exp. Psych.*, Vol. 70 (1965), pp. 394-400.

Anderson, N. H. "Functional Measurement and Psychophysical Judgments," *Psychol. Rev.*, Vol. 77 (1970), pp. 63-65.

Anderson, N. H. "Looking for Configurality in Clinical Judgment," *Psychol. Bull.*, Vol. 78 (1972), pp. 93-102.

Anderson, N. H. and J. C. Shanteau. "Information Integration in Risky Decision-Making," *J. Exp. Psych.*, Vol. 84 (1970), pp. 441-451.

Anderson, N. H. and J. Shanteau. "Weak Interferences With Linear Models," *Psychol. Bull.*, Vol. 84 (1977), 1155-1170.

Analytics. *Measuring the Performance of Operational Decision Aids (Final Report 1161-B)*. Willow Grove, PA, Analytics, 1976.

Bar-Hillel, M. "The Base Rare Fallacy in Probability Judgements." Decision Research, Eugene, Ore., 1977. (Decision Research Report No. 77-4.)

Beach, B. H. "Expert Judgment About Uncertainty: Bayesian Decision-Making in Realistic Settings," in *Organiz. Behav. and Hum. Perform.*, Volume 14 (1975), pp. 10-59.

Beach, L. R. "Probability, Magnitudes and Conservative Revision of Subjective Probabilities," *J. Exp. Psych.*, Vol. 77 (1969), pp. 57-63.

Beals, R., D. H. Krantz, and A. Tversky. "Foundations of Multidimensional Scaling," *Psychol. Rev.*, Vol. 75 (1968), pp. 127-144.

Becker, G. M., M. H. DeGroot, and J. Marschak. "Measuring Utility by a Single-Response Sequential Method," *Behav. Sci.*, Vol. 9 (1964), pp. 226-232.

Brehmer, B. "Hypotheses About Relations Between Scaled Variables in the Learning of Probabilistic Inference Tasks," in *Organiz. Behav. and Hum. Perform.*, Vol. 11 (1974), pp. 1-27.

Brehmer, B., J. Kuylenstierna, and J. Liljergren. "Effects of Function Form and Cue Validity on Subjects' Hypotheses in Probabilistic Inference Tasks," in *Organiz. Behav. and Hum. Perform.*, Vol. 11 (1974), pp. 338-354.

Brehmer, B., and G. Quarnstrom. "Information Integration and Subjective Weights in Multiple-Cue Judgments," in *Organiz. Behav. and Hum. Perform.*, Vol. 17 (1976), pp. 118-126.

Brickman, P., and S. M. Pierce. "Estimates of Conditional Probabilities of Confirming Versus Disconfirming Events as a Function of Inference Situation and Prior Evidence," in *J. Exp. Psych.*, Vol. 95 (1972), pp. 235-237.

Briggs, G. E. "On the Predictor Variable for Choice Reaction Time," Memory and Cognition, 1974.

Briggs, G. E., and J. M. Swanson. "Encoding, Decoding and Control Functions in Human Information Processing," *J. Exp. Psych.*, Vol. 86 (1970), pp. 296-308.

Brown, R. V., A. S. Kahr, and C. Peterson. *Decision Analysis: An Overview*. New York, Holt, Rinehart, and Winston, 1974. 85 pp.

Brown, T. A., and E. H. Shuford. *Quantifying Uncertainty Into Numerical Probabilities for the Reporting of Intelligence*. Santa Monica, CA, Rand Corporation, July 1973. 47 pp.

Castellan, J. J., Jr. "The Analysis of Multiple Criteria in Multiple-Cue Judgment Tasks," in *Organiz. Behav. and Hum. Perform.*, Vol. 8 (1972), pp. 242-261.

Castellan, J. J., Jr. "Comments on the 'Lens Model' Equation and the Analysis of Multiple-Cue Judgment Tasks," in *Psychometrika*, Vol. 38 (1973), pp. 87-100.

Chinnis, J. O., and C. R. Peterson. "Inferences About a Nonstationary Process," *J. Exp. Psych.*, Vol. 77 (1968), pp. 620-625.

Chinnis, J. O., and C. R. Peterson. "Nonstationary Processes and Conservative Inferences," *J. Exp. Psych.*, Vol. 84 (1970), pp. 248-251.

Coombs, C. H. "On the Use Inconsistency of Preferences in Psychological Measurement," *J. Exp. Psych.*, Vol. 55 (1958), pp. 511-518.

Coombs, C. H., R. M. Dawes, and A. Tversky. *Mathematical Psychology*, Englewood Cliffs, N. J., Prentice-Hall, Inc., 1970. 440 pp.

Coombs, C. H., and S. S. Komorita. "Measuring the Utility of Money Through Decisions," *Amer. J. Psychol.*, Vol. 71 (1958), pp. 383-89.

Coombs, C. H., and D. G. Pruitt. "Components of Risk in Decision-Making: Probability and Variance Preferences," *J. Exp. Psych.*, Vol. 60 (1960), pp. 265-77.

Crooks, Kuppin, and Freddy. "Applications of Adaptive Decision Aiding Systems to Computer-Assisted Instruction: Adaptive Computerized Training System (ACTS)." U.S. Army Research Institute, Arlington, Va., 1977. (TR-PRTR-1028-77-1.)

Dawes, R. M. "Predictive Models as a Guide to Preference," in *Proceedings of the 1975 International Conference on Cybernetics and Society*, San Francisco, CA, 1975.

Dawes, R. M., and B. Corrigan. "Linear Models in Decision-Making," *Psychol. Bull.*, Vol. 81 (1974), pp. 95-106.

Decision-Science Applications, Inc. "An Emissions Control Decision Aid," Arlington, Va., 1978. (Report No. DSA-66.)

DeFinetti, B. "Logical Foundations and Measurement of Subjective Probability," *Acta Psychologica*, Vol. 34 (1970), pp. 129-145.

Domas, P. A., and C. R. Peterson. "Probabilistic Information Processing Systems: Evaluation With Conditionally Dependent Data," *Organiz. Behav. and Hum. Perform.*, Vol. 7 (1972), pp. 77-85.

Drapper, N. R., and H. Smith. *Applied Regression Analysis*. New York, John Wiley & Sons, Inc., 1966, 411 pp.

DuCharme, W. M. "A Response Bias Explanation of Conservative Human Inference," *J. Exp. Psych.*, Vol. 85 (1970), pp. 66-75.

DuCharme, W. M. and C. R. Peterson. "Intuitive Inferences About Normally Distributed Populations," *J. Exp. Psych.*, Vol. 78 (1968), pp. 269-75.

Eckenrode, R. T. "Weighting Multiple Criteria," *Manage. Sci.*, Vol. 12 (1965), pp. 180-91.

Edwards, W. "The Prediction of Decisions Among Bets," *J. Exp. Psych.*, Vol. 50 (1955), pp. 201-10.

Edwards, W. *Computer-Aided Decision Making: A Short Course*, Class Lecture, UCLA, Spring, 1978.

Edwards, W. H. Lindman, and L. J. Savage. "Bayesian Statistical Inference for Psychological Research," *Psychol. Rev.*, Vol. 70 (1963), pp. 193-242.

Edwards, W., and P. Slovic. "Seeking Information To Reduce the Risk of Decisions," *Amer. J. Psychol.*, Vol. 78 (1965), pp. 188-197.

Eils, L. C., D. A. Seaver, and W. Edwards. "Developing the Technology of Probabilistic Inference: Aggregating by Averaging Reduces Conservatism." University of Southern California, Los Angeles, Calif., 1977. (SSRI Research Report No. 77-3.)

Feigenbaum, E. A. "Computer-Aided Decision Making: A Short Course," class lecture, UCLA, spring 1978.

Fitts, P. M. "The Information Capacity of the Human Motor System in Controlling Amplitude of Movement," *J. Exp. Psych.*, Vol. 47 (1954), pp. 381-91.

Fitts, P. M. "Cognitive Aspects of Information Processing: III Set for Speed Versus Accuracy," *J. Exp. Psych.*, Vol. 71 (1966), pp. 849-57.

Fischer, G. W. "Experimental Aplications of Multi-Attribute Utility Models," in *Utility, Probability, and Human Decision Making*, ed. D. Wendt, C. A. J. Vlek, pp. 7-46. Dordrecht, The Netherlands, Reidel, 1975. 418 pp.

Fischer, G. W. "Multidimensional Utility Models for Risky and Riskless Choice," *Organiz. Behav. and Hum. Perform.*, Vol. 17 (1976), pp. 127-46.

Fischhoff, B. "Hindsight ≠ Foresight: The Effect of Outcome Knowledge on Judgment Under Uncertainty," *J. Exp. Psych.: Human Perception and Performance*, Vol. 1 (1975), pp. 288-99.

Fischhoff, B. "Perceived Informativeness of Facts," *J. Exp. Psych.*, Vol. 3 (1977), pp. 349-351.

Fischhoff, B. "Informal Use of Formal Models," presented, First Annual Meeting of the Institute of Management Science and Operation Research Society of America, May 1978.

Fischhoff, B. "Clinical Decision Analysis," *Operations Research*, in press.

Fischhoff, B., Slovic P., and Lichtenstein, S. "Knowing What You Want, Measuring Labile Values" in *Cognitive Processes in Choice and Decision Behavior*, ed. T. Wallsten, Hillsdale, N.J.: Lawrence Erlbaum Associates, Inc. In press.

Fischhoff, B., P. Slovic, and S. Lichtenstein. "Knowing With Certainty the Appropriateness of Extreme Confidence," *J. Exp. Psych.: Human Perception and Performance*, in press.

Galanter, R. "The Direct Measurement of Utility and Subjective Probability," *Amer. J. Psych.*, Vol. 75 (1962), pp. 208-20.

Gardiner, P. C. and W. Edwards. "Public Values: Multi-Attribute Utility Measurement for Social Decisions," in *Proceedings of the 1975 International Conference on Cybernetics and Society*, San Francisco, CA.

Gettys, C. F., S. D. Fisher, and T. Mehle. *Hypothesis Generation and Plausibility Assessment.* University of Oklahoma, Norman, 1978. (Annual Report TR 15-10-78.)

Gettys, C. F., C. W. Kelly, III, and C. R. Peterson. "The Best Guess Hypothesis in Multistage Inference," *Organiz. Behav. and Hum. Perform.*, Vol. 10 (1973), pp. 364-73.

Godden, D. R. "Probability as Preparedness To Act: Direct Estimation Versus Indirect Measurement," *Organiz. Behav. and Hum. Perform.*, Vol. 17 (1976), pp. 147-58.

Goldberg, L. R. "Man vs. Model of Man: A Rationale, Plus Some Evidence, for a Method of Improving on Clinical Inferences," *Psychol. Bull.*, 1970, Vol. 73, pp. 422-32.

Graesser, C. C., and N. H. Anderson. "Cognitive Algebra of the Equation: Gift Size = Generosity X Income," *J. Exp. Psych.*, Vol. 103 (1974), pp. 692-99.

Green, D. M., and J. A. Swets. *Signal Detection Theory and Psychophysics.* New York, John Wiley and Sons, Inc., 1966. 449 pp.

Hammond, K. R., and D. A. Summers. "Cognitive Control," *Psychol. Rev.*, Vol. 79 (1972), pp. 58-67.

Hammond, K. R., T. R. Stewart, B. Brehmer, and D. O. Steinmann. "Social Judgment Theory" in *Human Judgment and Decision Processes.* New York, Academic Press, 1975. Pp. 271-312.

Hamner, W. C., and P. L. Carter. "A Comparison of Alternate Production Management Coefficient Decision Rules," *Decision Sci.*, Vol. 6 (1975), pp. 324-36.

Hoel, P. G. *Mathematical Statistics.* New York, John Wiley and Sons, Inc., 1971, 402 pp.

Hoepfl, R., and G. Huber. "A Study of Self-Explicated Utility Models," *Behav. Sci.*, Vol. 15 (1970), pp. 408-14.

Hoffman, J., and C. Peterson. *A Scoring Rule To Train Probability Assessors.* University of Michigan, Engineering Psychology Laboratory Technical Report 037230-4-T, September 1972.

Huber, G. "Multi-Attribute Utility Models: A Review of Field and Field-Like Studies," *Manage. Sci.*, Vol. 20 (1974), pp. 1393-1402.

Huber, G., R. Daneshaar, and D. Ford. "An Empirical Comparison of Five Utility Models for Predicting Job Preferences," *Organiz. Behav. and Hum. Perform.*, Vol. 6 (1971), pp. 267-82.

Humphreys, P., and A. Humphreys. "An Investigation of Subjective Preference Orderings for Multi-Attributed Alternatives," in *Utility, Probability, and Human Decision Making*, ed. D. Wendt and C. J. Vlek, Dordrecht, The Netherlands, Reidel, 1975. 481 pp.

Integrated Sciences Corporation. *Experimental Investigation of the Usefulness of Operator Aided Optimization in a Simulated Tactical Decision Aiding Task.* Santa Monica, Calif., 1978. (Report No. 215-4.)

Johnson, E. M., R. C. Cavangah, R. L. Spooner, and M. G. Samet. "Utilization of Reliability Measurement in Bayesian Inference: Models and Human Performance," *Inst. Elec. Electron. Eng., Trans. on Reliability*, Vol. R22 (1973), pp. 176-83.

Kahneman, D. and A. Tversky. "Subjective Probability: A Judgment of Representativeness," *Cognitive Psych.*, Vol. 3 (1972), pp. 430-54.

Kahneman, D., and A. Tversky. "On the Psychology of Prediction," *Psychol. Rev.*, Vol. 80 (1973), pp. 237-51.

Kahneman, D., and A. Tversky. "Prospect Theory: An Analysis of Decision Under Risk," *Econometrica*, in press.

Kahneman, D., and A. Tversky. "Intuitive Prediction: Biases and Corrective Procedures," in *Management Science*, [in press].

Keeney, R. L., and H. Raiffa. *Decisions With Multiple Objectives: Preferences and Value Tradeoffs.* New York, John Wiley and Sons, 1976. 549 pp.

Keeney, R. L., and A. Skherman. "Assessing and Analyzing Preferences Concerning Multiple Objectives: An Interactive Computer Program." *Behav. Sci.*, Vol. 21 (1976), pp. 173-83.

Klahr, D. "Decision Making in a Complex Environment," *Manage. Sci.*, Vol. 13 (1969), pp. 595-618.

Kneppreth, N. P., D. H. Gustafson, R. P. Leifer, and F. M. Johnson. "Techniques for the Assessment of Worth," *Tech. Paper 254*, published at the Army Research Institute, Arlington, VA, 1974.

Krantz, D. H. R. D. Luce, P. Suppes, and A. Tversky. *Foundations of Measurement: Additive and Polynomial Representations, Volume 1.* New York, Academic Press, 1971. 560 pp.

Leal, A., S. Levin, S. Johnston, M. Agmon, and G. Weltman. *An Interactive Computer Aiding System for Group Decision Making.* Perceptronics, Woodland Hills, CA., Technical Report PQTR-1046-78-2, February 1978.

Leal, A., and J. Pearl. "An Interactive Program for Conversational Elicitation of Decision Structures." *Proceedings of the IEEE Conference on Systems, Man and Cybernetics*, May 1977, Vol. SMC-7(5), pp. 368-76.

Lee, W. *Decision Theory and Human Behavior.* New York, John Wiley & Sons, Inc., 1971. 345 pp.

Levine, J. M., and M. G. Samet. "Information Seeking With Multiple Sources of Conflicting and Unreliable Information," *Human Factors*, Vol. 15 (1973), pp. 407-19.

Levine, J. M., M. G. Samet, and R. E. Branhlek. "Information Seeking With Limitations on Available Information and Resources," *Human Factors*, Vol. 17 (1975), pp. 502-13.

Lichtenstein, S. "Bases for Preferences Among Three-Outcome Bets," *J. Exp. Psych.*, Vol. 69 (1965), pp. 162-69.

Lichtenstein, S. "Conditional Non-Independence of Data in a Practical Bayesian Decision Task," *Organiz. Behav. and Hum. Perform.*, Vol. 8 (1972), pp. 21-25.

Lichtenstein, S., and B. Fischhoff. "Do Those Who Know More Also Know More About How Much They Know?" *Organiz. Behav. and Hum. Perform.*, Vol. 20 (1977), pp. 159-183.

Lichtenstein, S., and B. Fishhoff. *Training for Calibration.* Decision Research, Eugene, Ore., 1978. (Report PTR 1047-78-5.)

Lichtenstein, S., B. Fischhoff, and L. Phillips. "Calibration of Probabilities: The State of the Art," in *Decision Making and Change in Human Affairs*, ed. H. Jungermann and G. deZeeuw, Dordrecht, The Netherlands, D. Reidel Publishing Co., 1977. pp. 275-324.

Linolcy, D. V., A. Tversky, and R. V. Brown. *On the Reconciliation of Probability Assessments*. Decision and Designs, Inc., McLean, Va., 1977. (Report TR 77-10-25.)

Luce, D., and E. Shipley. "Probability Gambles as a Step Function of Event Probability," *J. Exp. Psych.*, Vol. 63 (1962), pp. 42-49.

Luce, D., and P. Suppes. "Preference, Utility and Subjective Probability," in *Handbook of Mathematical Psychology*, Vol. III, ed. D. Luce, R. R. Bush, and E. Galanter. New York, John Wiley & Sons, Inc., 1965. p. 514.

Lyons, D., and P. Slovic. "Dominance of Accuracy Information and Neglect of Base Rates in Probability Estimation," *Acta Psychologica*, Vol. 40 (1976), pp. 287-98.

MacCrimmon, K. R. "Decision-Making Among Multiple-Attribute Alternatives: A Survey and Consolidated Approach," *Memorandum RM-4823-ARPA*, December 1968, Santa Monica, CA., The Rand Corporation, 60 pp.

March, J. G. "Bounded Rationality, Ambiguity, and the Engineering of Choice," *Bell Journal of Economics*, 1978, pp. 587-608.

Marks, D. F., and J. K. Clarkson. "An Explanation of Conservatism in the Bookbag-and-Pokerchips Situation," *Acta Psychologica*, Vol. 36 (1972), pp. 145-60.

Marks, D. F., and J. K. Clarkson. "Conservatism as Non-Bayesian Performance: A Reply to DeSwart," *Acta Psychologica*, Vol. 37 (1973), pp. 55-63.

Matheson, J. E. "Decision Analysis: Examples and Insights," presented at the Fifth International Conference on Operational Research, Venice, Italy, 1969.

McClelland, G. *Equal Versus Differential Weighting for Multiattribute Decisions: There are No Free Lunches*. Center for Research on Judgment and Policy, University of Colorado, Boulder, CO, 1978, Report No. 207, 31 pp.

Miller, G. A. "The Magical Number Seven Plus or Minus Two: Some Limits on Our Ability To Process Information," *Psychol. Rev.*, Vol. 63 (1956), pp. 81-97.

Murphy, A. H., and R. L. Winkler. "Scoring Rules in Probability Assessment and Evaluation," *Acta Psychologica*, Vol. 34 (1970), pp. 273-86.

Murphy, A. H., and R. L. Winkler. "Forecasters and Probability Forecasts: Some Current Problems," *Amer. Meteorol. Soc., Bull.*, Vol. 52 (1970), pp. 234-47.

Murphy, A. H., and R. L. Winkler. "Subjective Probability Forecasting: Some Experiments in the Real World," in *Utility, Probability, and Human Decision Making*, ed. D. Wendt and C. Vlek. Dordrecht, The Netherlands, D. Reidel, 1975. 403 pp.

Nickerson, R. S., and G. E. Feehrer. *Decision Making and Training*. Naval Training Equipment Center, Orlando, FL, 1975. (Technical Report 73-C-0128-1, 210 pp.)

Nisbett, R. E., and E. Borgida. "Attribution and the Psychology of Prediction," *J. Personality and Social Psychology*, Vol. 32 (1975), pp. 932-43.

Ölander, F. "Search Behavior in Non-Simultaneous Choice Situations: Satisficing or Maximizing," in *Utility, Probability and Human Decision Making*, ed. D. Wendt and C. Vlek. Dordrecht, The Netherlands, Reidel, 1975. 417 pp.

Pearl, J. "A Framework for Processing Value Judgments," *IEEE Trans. on Systems, Man, and Cybernetics*, Vol. SMC-7, (1977), pp. 349-54.

Pachella, R. G., and R. W. Pew. "Speed-Accuracy Tradeoff in Reaction Time: Effect of Discrete Criterion Times," *J. Exp. Psych.*, Vol. 76 (1968), pp. 19-24.

Perceptronics. *Adaptive Computer Aiding in Dynamic Decision Processes*. Woodland Hills, CA, August 1976. (Final Technical Report PFTR-1016-76-8/30.)

Peterson, C. R., and L. R. Beach. "Man as an Intuitive Statistician," *Psychol. Bull.*, Vol. 68 (1967), pp. 29-46.

Peterson, C. R., N. M. DuCharme, and W. Edwards. "Sampling Distributions and Probability Revisions," *J. Exp. Psych.*, Vol. 76 (1968), pp. 236-43.

Pew, R. W. "Speed-Accuracy Operating Characteristic," *Acta Psychologica*, Vol. 30 (1969), pp. 19-24.

Pitz, G. "Information Seeking When Available Information Is Limited," *J. Exp. Psych.*, Vol. 76 (1968), pp. 25-34.

Pitz, G. "A Structural Theory of Uncertain Knowledge," in *Utility Probability, and Human Decision Making*, ed. D. Wendt and C. Vlek. Dordrecht, The Netherlands, Reidel, 1975, 417 pp.

Pratt, J. W. "Risk Aversion in the Small and in the Large," *Econometrica*, Vol. 32 (1964), pp. 122-37.

Raiffa, H. *Decision Analysis: Introductory Lectures on Choices Under Uncertainty*. Reading, Mass., Addison-Wesley, 1968. 300 pp.

Rapoport, A. "Research Paradigms for Studying Dynamic Decision Behavior," in *Utility, Probability, and Human Decision Making*, ed. D. Wendt and C. Vlek. Dordrecht, The Netherlands, D. Reidel, 1975. 407 pp.

Royden, H. L., P. Suppes, and K. Walsh. "A Model for the Experimental Measurement of the Utility of Gambling," *Behav. Sci.*, Vol. 4 (1959), pp. 11-18.

Saleh, J., A. Leal, L. Lucaccini, P. Gardiner, and R. Hopf-Weichel. *Analysis of Requirements and Methodology for Decision Training in Operational Systems*. Naval Training Equipment Center, Orlando, Fla., 1978. (NAVTRANEQUIPCEN 77-0055-1.)

Savage, L. J. *The Foundations of Statistics*. New York, John Wiley & Sons, Inc., 1954. 282 pp.

Schum, D. A., I. L. Goldstein, and J. F. Southard. "Research on a Simulated Bayesian Information Processing System," *IEEE Trans. on Human Factors in Electronics*, Vol. HFE-7 (1966), pp. 37-48.

Schum, D. A., and P. E. Pfeiffer. "Observer Reliability and Human Inference," *IEEE Trans. on Reliability*, Vol. R-22 (1973), pp. 170-75.

Schaeffer, F. E. "The Evaluation of Individual and Aggregated Subjective Probability Distributions," *Organiz. Behav. and Hum. Perform.*, Vol. 17 (1976), pp. 199-210.

Shanteau, J. H. "Descriptive Versus Normative Models of Sequential Inference Judgment," *J. Exp. Psychol.*, Vol. 83 (1972), pp. 63-68.

Shanteau, J. "Component Processes in Risky Decision Making," *J. Exp. Psych.*, Vol. 103 (1974), pp. 680-90.

Sheriden, J. E., M. D. Richards, and J. W. Slocum. "Comparative Analysis of Expectancy and Heuristic Models of Decision Behavior," *J. Appl. Psych.*, Vol. 60 (1975), pp. 361-80.

Shortliffe, E. H., and B. G. Buchanan. "A Model of Inexact Reasoning in Medicine," *Math. Biosci.*, Vol. 23 (1975), pp. 351-379.

Simon, H. A. *The Sciences of the Artificial*. Cambridge, Mass., M.I.T. Press, 1969, 123 pp.

Sinaiko, H. W. *Operational Decision Aids: A Program of Applied Research for Naval Command and Control Systems*. Smithsonian Institution, Washington, D.C., 1977. (TR-5.)

Slovic, P. "Value as a Determiner of Subjective Probability," *IEEE Trans. on Human Factors in Electronics*, Vol. HFE-7 (1966), pp. 22-28.

Slovic, P. *From Shakespeare to Simon: Speculations--and Some Evidence--About Man's Ability To Process Information.* Oregon Research Institute, Eugene, OR, 1972. (ORI Research Monograph, Vol. 12, 37 pp.)

Slovic, P. "Towards Understanding and Improving Decisions," ed. E. A. Fleischman, in *Human Performance and Productivity.* Decision Research, Inc., Eugene, Ore.

Slovic, P., personal communication, July 3, 1979.

Slovic, P., B. Fischhoff, and S. Lichtenstein. "Behavioral Decision Theory," *Annual Review of Psychology*, 1977, pp. 1-38.

Slovic, P., and S. Lichtenstein. "Probabilities and Payoffs in Risk Taking," *J. Exp. Psych. Monograph*, Vol. 78 (1968), pp. 1-17.

Slovic, P., and S. Lichtenstein. "Comparison of Bayesian and Regression Approaches to the Study of Information Processing in Judgment," *Organiz. Behav. and Hum. Perform.*, Vol. 6 (1971), pp. 649-743.

Slovic, P., and A. Tversky. "Who Accepts Savage's Axioms?" *Behav. Sci.*, Vol. 19 (1975), pp. 368-73.

Snapper, K. J., and P. Fryback. "Inferences Based on Unreliable Reports," *J. Exp. Psychol.*, Vol. 81 (1971), pp. 401-404.

Snapper, K. J., and C. R. Peterson. "Information Seeking and Data Diagnosticity," *J. Exp. Psych.*, Vol. 87 (1971), pp. 429-33.

Stäel von Holstein, C. A. "Probabilistic Forecasting: An Experiment Related to the Stock Market," *Organiz. Behav. and Hum. Perform.*, Vol. 8 (1972), pp. 139-158.

Sternberg, S. "Memory-Scanning: Mental Processes Revealed by Reaction Time Experiments," *Amer. Scientist*, Vol. 57 (1969), pp. 421-57.

Suppes, P., and K. Walsh. "A Non-Linear Model for the Measurement of the Utility of Gambling," *Behav. Sci.*, Vol. 4 (1959), pp. 204-11.

Swensson, R. G., and W. Edwards. "Response Strategies in a Two-Choice Reaction Task With Continuous Cost for Time," *J. Exp. Psych.*, Vol. 88 (1971), pp. 67-81.

Tversky, A. "Utility Theory and Additivity Analysis of Risky Choices," *J. Exp. Psych.*, Vol. 75 (1967), pp. 175-202.

Tversky, A. "Intransitivity of Preferences," *Psychol. Rev.*, Vol. 76 (1969), pp. 31-48.

Tversky, A. "Choice by Elimination," *J. Math. Psych.*, Vol. 9 (1972a), pp. 341-67.

Tversky, A. "Elimination by Aspect: A Theory of Choice," *Psychol. Rev.*, Vol. 79 (1972b), pp. 281-99.

Tversky, A., and D. Kahneman. "Judgment Under Uncertainty: Heuristics and Biases," in *Utility, Probability and Human Decision Making*, ed. D. Wendt and C. Vlek. Dordrecht, The Netherlands, Reidel, 1975. 417 pp.

Tversky, A. and D. A. Kahneman. A Causal Schemata in Judgments Under Uncertainty. Decision Research, Eugene, Ore., 1977. (TR PTR-1060-77-10.)

Vlek, C. B. "Response Modes of Probability Estimation," *Organiz. Behav. and Hum. Perform.*, Vol. 9 (1973), pp. 460-81.

VonWinterfeldt, D., and G. W. Fischer. "Multiattribute Utility Theory: Models and Assessment Procedures," in *Utility, Probability, and Human Decision Making*, ed. C. Vlek and D. Wendt. Dordrecht, The Netherlands, Reidel, 1975. 417 pp.

Wendt, D. "Value of Information for Decisions," *J. Math. Psych.*, Vol. 6 (1969), pp. 430-43.

Winston, P. H. *Artificial Intelligence*. Reading, Mass., Addison-Wesley, Inc. (1977), 441 pp.

Winkler, R. L., and A. H. Murphy. "Experiments in the Real World and in the Laboratory," *Organiz. Behav. and Hum. Perform.*, Vol. 10 (1973), pp. 252-70.

Weyer, R. S. "Information Redundancy, Inconsistency and Novelty and Their Role in Impression Formation," *J. Exp. Social Psych.*, Vol. 6 (1970), pp. 111-27.

Yelloit, J. I., Jr. "Correction for Fast Guessing and the Speed-Accuracy Tradeoff in Choice Reaction Time," *J. Math. Psych.*, Vol. 8 (1971), pp. 159-99.

Zlonick, J. "A Theorem for Prediction," *Foreign Service Journal*, Vol. 45 (1968), pp. 20-41.

INITIAL DISTRIBUTION

6 Naval Air Systems Command
  AIR-04 (1)
  AIR-340F (1)
  AIR-5313 (2)
  AIR-954 (2)
1 Chief of Naval Operations (OP-987)
1 Chief of Naval Material (MAT-0344)
3 Naval Sea Systems Command
  SEA-03416 (1)
  SEA-09G32 (2)
1 Chief of Naval Research, Arlington
1 Bureau of Medicine & Surgery (Code 513)
1 Naval Aerospace Medical Research Laboratory, Pensacola (Code L5)
4 Naval Air Development Center, Warminster
  Code 602 (1)
  Code 6021 (1)
  Code 6022 (1)
  Technical Library (1)
1 Naval Air Test Center (CT-176), Patuxent River (SY-72)
1 Naval Postgraduate School, Monterey (Technical Library)
1 Naval Training Equipment Center, Orlando (Technical Library)
2 Navy Personnel Research and Development Center, San Diego
  Code 02 (1)
  Code 03 (1)
1 Office of Naval Research Branch Office, Pasadena
2 Pacific Missile Test Center, Point Mugu
  Code 1226 (1)
  Technical Library (1)
1 Army Human Engineering Laboratory, Aberdeen Proving Ground
1 Aerospace Medical Research Laboratory, Wright-Patterson Air Force Base (Code HEA)
12 Defense Technical Information Center
1 Defense Intelligence Agency
1 Ames Research Center (NASA) (Aviation Safety Research Office, Technical Library)
1 Applied Physics Laboratory, JHU, Laurel, MD
1 Hughes Aircraft Company, Culver City, CA (Display Systems Laboratory)
1 Human Factors Research, Incorporated, Goleta, CA
1 Institute for Defense Analyses, Arlington, VA (Technical Library)
2 McDonnell Douglas Corporation, St. Louis, MO (Engineering Psychology)
1 Perceptronics, Woodland Hills, CA (Technical Library)
1 The Boeing Company, Seattle, WA (Crew Systems MS-41-44)
1 The Rand Corporation, Santa Monica, CA (Library)